

The Journal of Educational Psychology

*Devoted Primarily to the Scientific Study of Problems of
Learning and Teaching*

BOARD OF EDITORS*

WM. CLARK TROW, *Chairman*
University of Michigan

STEPHEN M. COREY
Teachers College
Columbia University

J. B. STROUD
State University of Iowa

*Appointed by the American Psychological Association

EDITORIAL CONSULTANTS

PAUL BLOMMERS
CHESTER W. HARRIS
HAROLD E. JONES

H. H. REMMERS
PERCIVAL M. SYMONDS

MILES A. TINKER
ALEXANDER G. WESMAN
PAUL A. WITTY



Baltimore, Md.
WARWICK and YORK, Inc.
1954

TABLE OF CONTENTS

Applicational Transfer and Inhibition. DONALD E. P. SMITH.....	169-174
Attitudes of Parents of Differing Socio-economic Status toward the Teaching Profession. WILLIAM F. ANDERSON, JR.....	345-352
Book Reviews. 60-64, 121-128, 182-192, 250-254, 310-317, 372-384, 436-444	
Classroom Environment and Pupil Welfare. NEWELL C. KEPHART AND WILLIAM FLOYD.....	52-53
College Grades and Self-estimates of Intelligence. ORVILLE G. BRIM, JR.....	477-484
Comparative Decline of Graduate Record Examination Scores and Intelligence with Age. R. TRAVIS OSBORNE AND WILMA B. SANDERS.....	353-358
Comparison of Mental Abilities of Bright and Dull Children of Comparable Mental Ages, A. OLIVER P. KOLSTOE.....	161-168
Comparison of the Stanford-Binet, Revised Form L, and the California Test of Mental Maturity (S Form). WILLIAM D. SHELDON AND GEORGE MANOLAKES...	499-504
Comparison of the Value Systems of Southern Negro and Northern White Youth, A. JULIUS H. HUGHES AND GEORGE C. THOMPSON.....	300-309
Delineating the Areas of Educational and Clinical Psychology. RUSSELL N. CASSEL.....	292-299
Determining Students' Concepts of Effective Teaching from Their Ratings of Instructors. WILLIAM E. COFFMAN.....	277-286
Differential Responses among College Students on the Vocabulary Subtest of the Wechsler Intelligence Scale. CHALMERS L. STACEY AND S. WILLIAM SPANIER.....	29-35
Educational Therapy. BRUCE McCARTY.....	65-80
Effect of Slanted Text upon the Readability of Print. MILES A. TINKER.....	287-291
Experiment in Ninth-grade Reading Improvement, An. ROBERT F. BARRY AND PAUL E. SMITH.....	407-414
Experimental Comparison of Recitation, Discussion, and	

Tutorial Methods in College Teaching, An. HAROLD GUETZKOW, E. LOWELL KELLY AND W. J. McKEACHIE.....	193-207
Experimental Study of the Reading-rate Controller, An. GEORGE F. WOOSTER.....	421-426
Factor Analysis of Student Ratings of Psychology Instructors on the Purdue Scale, A. A. W. BENDIG..	385-393
Falsification of Interest Patterns on the Kuder Preference Record. EDWARD J. DURNALL, JR.....	240-243
Final Examination Performance in Traditional and Cross-disciplinary Courses. IRWIN A. BERG AND RAYMOND J. MURPHY.....	365-371
Group Procedures Modifying Attitudes of Prejudice in the College Classroom. GERALD S. WIEDER.....	332-344
Group-discussion Method of Teaching Psychology, A. ALBERT EGLASH.....	257-267
Hostility: An Educational Paradox. HARRY A. GRACE.	432-435
How Student-teachers Identify Responsibility in Children. DALE B. HARRIS.....	233-239
Improved Formula for Scoring Certain Guess-who Ratings at the Adolescent Level, An. EVAN R. KEISLAR.....	151-160
✓ Influence of Social Class upon Performance on the Draw-a-Man Test. JOSEPH H. BRITTON.....	44-51
Item Validity and Response Change under Two Different Testing Conditions. DOUGLAS G. SCHULTZ.....	36-43
✓ Motivational Differences between High and Low Scholarship College Students. WILLIAM F. BROWN, NORMAN ABELES AND IRA ISCOE.....	215-223
Perception in Number Skills—A Study in Tachistoscopic Training. JOHN L. PHILLIPS, JR.....	459-468
Personal Values and the Emphatic Responses: Their Inter-relationships. GLENN R. HAWKES AND ROBERT L. EGBERT.....	469-476
Problem-oriented Approach to Teaching Psychology, The. W. J. McKEACHIE AND WESLEY HILER.....	224-232
Psychological Prerequisites to Educational Psychology, The. ROY GLADSTONE.....	415-420
Publications Received.....	255-256, 318-320, 445-448
Reading Abilities of Bright and Dull Children of Comparable Mental Ages. EMERY P. BLIESMER.....	321-331

Relationship between Class Absences and College Grades, The. ALEXIS M. ANIKEEFF.....	244-249
Relationships between Eight Variables and F-Test Scores of Teachers. H. H. REMMERS AND MARTIN STEIN- BERG.....	427-431
Response Strength in a Classroom Task Related to a 'Forward' Delay in Reinforcement. DONAVON AUBLE AND EDMUND MECH.....	175-181
Rorschach Configurations Associated with College Achievement. CHARLES C. MCARTHUR AND STAN- LEY KING.....	492-498
Some Increasing Parent-Child Similarities during the Growth of Children. NANCY BAYLEY.....	1-21
Sophomore General Culture Tests and Student Knowl- edge of Cultural Facts, The. BERT D. ANDERSON..	359-364
Spelling Ability and Vocabulary Level of One Hundred College Freshmen. FRANK LAYCOCK.....	485-491
Student Attitude and Achievement in a Course in Intro- ductory Statistics. A. W. BENDIG AND J. B. HUGHES, III.....	268-276
Student-centered versus Instructor-centered Instruction. W. J. McKEACHIE.....	143-150
Student Differences in Competency in Educational Psy- chology Prior to Formal Training. WILLIAM W. LYNCH, JR.....	110-120
Study of Transfer, A: Sex Differences in the Reasoning Process. MAX MARTIN KOSTICK.....	449-458
Topographical Orientation in Naval Aviation Cadets? BRANT CLARK AND R. DANIEL MALONE.....	91-109
Two Methods of Teaching English Composition to Re- tarded College Freshmen. RAY CHARLES MAIZE..	22-28
Unreliability of Group Test Profiles. FLORENCE T. SNODGRASS.....	129-142
Validity of an Objective Examination for the Placement of Foreign Students in English Courses. IRVING LORGE AND LORRAINE KRUGLOV DIAMOND.....	208-214
Validity of Teachers' Ratings of Adolescents' Adjustment and Aspirations. DAVID P. AUSUBEL, HERBERT M. SCHIFF, AND MAJORIE P. ZELENY.....	394-406
Who Is Penalized for the Penalty for Guessing? ALEX C. SHERRIFFS AND DONALD S. BOOMER.....	81-90

BOOK REVIEWS

Roger G. Barker, Beatrice A. Wright, Lee Meyerson, and Mollie R. Gonick. Adjustment to Physical Handicap and Illness: A Survey of the Social Psychology of Physique and Disability. (H. MELTZER).....	60
Leopold Bellak, Editor. Psychology of Physical Illness. Psychiatry Applied to Medicine, Surgery, and the Specialties. (WILLIAM SCHOFIELD).....	380
M. E. Bennett. College and Life. (CECIL M. FREEBURNE).....	378
Harold W. Bernard. Mental Hygiene for Classroom Teachers. (A. S. EDWARDS).....	183
Oscar Krisen Buros, Editor. The Fourth Mental Measurements Yearbook. (LEE J. CRONBACH).....	314
Drug Addiction among Adolescents. (H. MELTZER).....	377
Ernest Gardner. Fundamentals of Neurology. (H. MELTZER).....	127
Phyllis Greenacre. Trauma, Growth and Personality. (H. MELTZER).....	191
Bernard H. Hall, Mary Gangemi, V. L. Norris, Vivienne Hutchens Vail, and Gordon Sawatsky. Psychiatric Aide Education. (H. MELTZER).....	62
Charles Koch. The Tree Test. (GOLDINE C. GLESER).....	186
Oliver L. Lacey. Statistical Methods in Experimentation. (CECIL M. FREEBURNE).....	185
Robert Lindner. Prescription for Rebellion. (H. MELTZER).....	439
Frances Lloyd. Educating the Subnormal Child. (A. S. EDWARDS).....	443
Norman R. F. Maier. Principles of Human Relations. (DUANE C. SHAW).....	253
James L. Mursell. Using Your Mind Effectively. (PHILIP M. KITAY).....	188
C. W. Odell. How To Improve Classroom Testing. (CHESTER W. HARRIS).....	190
Charles E. Osgood. Method and Theory in Experimental Psychology. (A. S. EDWARDS).....	437
Leslie Phillips and Joseph G. Smith. Rorschach Interpretation: Advanced Technique. (LEE J. CRONBACH).....	125
Helen M. Robinson, Editor. Clinical Studies in Reading II. (MILES A. TINKER).....	182

Jerome M. Seidman, Editor. The Adolescent: A Book of Readings. (JOHN P. MCKEE).....	436
G. Wilson Shaffer and Richard S. Lazarus. Fundamental Concepts in Clinical Psychology. (A. S. EDWARDS)....	313
Everett L. Shostrom and Lawrence M. Brammer. The Dynamics of the Counseling Process. (LEONARD D. GOODSTEIN).....	250
B. F. Skinner. Science and Human Behavior. (PERCIVAL M. SYMONDS).....	121
Marjorie B. Smiley. Intergroup Education and the American College. (WILLIAM ELLER).....	2
William Stephenson. The Study of Behavior. Q-Technique and Its Methodology. (H. J. EYSENCK).....	374
Edwin Forbes Tait. Textbook of Refraction. (H. MELTZER)	63
John C. Townsend. Introduction to Experimental Method. (A. S. EDWARDS).....	443
Helen M. Walker and Joseph Lev. Statistical Inference. (FRANCIS G. CORNELL).....	372
Carl A. Whitaker and Thomas P. Malone. The Roots of Psychotherapy. (H. MELTZER).....	441
John W. M. Whiting and Irvin L. Child. Child Training and Personality. (LEONARD D. GOODSTEIN).....	310

INDEX TO AUTHORS

ABELES, NORMAN.....	215	KEPHART, NEWELL C.....	52
ANDERSON, BERT D.....	359	KING, STANLEY.....	492
ANDERSON, WILLIAM F., JR....	345	KOLSTOE, OLIVER P.....	161
ANIKEEFF, ALEXIS M.....	244	KOSTICK, MAX MARTIN.....	449
AUBEL, DONAVON.....	175	LAYCOCK, FRANK.....	485
ΔUSUBEL, DAVID P.....	394	LORGE, IRVING.....	208
BARRY, ROBERT F.....	407	LYNCH, WILLIAM W., JR.....	110
BAYLEY, NANCY.....	1	MAIZE, RAY CHARLES.....	22
BERG, IRWIN A.....	365	MALONE, DANIEL.....	91
BENDIG, A. W.....	268, 385	MANOLAKES, GEORGE.....	499
BLIESMER, EMERY P.....	321	MCCARTHUR, CHARLES C.....	492
BOOMER, DONALD S.....	81	MCCARTY, BRUCE.....	65
BRIM, ORVILLE G., JR.....	477	MCKEACHIE, W. J.....	143, 193, 224
BRITTON, JOSEPH H.....	44	MECH, EDMUND.....	175
BROWN, WILLIAM F.....	215	MURPHY, RAYMOND J.....	365
CASSEL, RUSSELL N.....	292	OSBORNE, R. TRAVIS.....	353
CLARK, BRANT.....	91	PHILLIPS, JOHN L., JR.....	459
COFFMAN, WILLIAM E.....	277	REMMERS, H. H.....	427
DIAMOND, LORRAINE KRUGLOV.	208	SANDERS, WILMA B.....	353
DURNALL, EDWARD J., JR.....	240	SCHIFF, HERBERT M.....	394
EGBERT, ROBERT L.....	469	SCHULTZ, DOUGLAS G.....	36
EGLASH, ALBERT.....	257	SHELDON, WILLIAM D.....	499
FLOYD, WILLIAM.....	52	SHERRIFFS, ALEX C.....	81
GLADSTONE, ROY.....	415	SMITH, DONALD E. P.....	169
GRACE, HARRY A.....	432	SMITH, PAUL E.....	407
GUETZKOW, HAROLD.....	193	SNODGRASS, FLORENCE T.....	129
HARRIS, DALE B.....	233	SPANIER, S. WILLIAM.....	29
HAWKES, ROBERT L.....	469	STACEY, CHALMERS L.....	29
HILER, WESLEY.....	224	STEINBERG, MARTIN.....	427
HUGHES, J. B., III.....	268	THOMPSON, GEORGE C.....	300
HUGHES, JULIUS H.....	300	TINKER, MILES A.....	287
ISCOE, IRA.....	215	WIEDER, GERALD S.....	332
KEISLAR, EVAN R.....	151	WOOSTER, GEORGE F.....	421
KELLY, E. LOWELL.....	193	ZELENY, MAJORIE P.....	394

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 45

January, 1954

Number 1

SOME INCREASING PARENT-CHILD SIMILARITIES DURING THE GROWTH OF CHILDREN

NANCY BAYLEY

Institute of Child Welfare, University of California

Age-changes in the correlation between children's mental test scores and some measure of parental ability have been shown for several different populations. Such reports have often been incidental to other aspects of research on intelligence. It has seemed worth while, therefore, to relate some of these findings to relevant material, in both mental and physical relationships, in the Berkeley Growth Study.

PARENT-CHILD CORRELATIONS IN THE LITERATURE

It is only rarely that mental test scores are available for the parents of tested children. But there are a few instances in which intelligence test scores of parents and their children have been compared directly. One of the most extensive and important of these was made by Jones and Conrad (8, 11, 12). They correlated Army Alpha test scores of parents with Alpha and Stanford-Binet scores of their children. In one hundred and forty families, in which both parents and children were given the Army Alpha test, four hundred and forty-one pairings gave a parent-child r of .49. The same r of .49 was obtained for five hundred and one pairings in one hundred eighteen families in which the children had the Stanford-Binet test. The r is also .49 for separate mother-child and father-child comparisons.¹

Among the most recently published parent-child test comparisons are the correlations presented by Skodak and Skeels between

¹ See (13) for a report on age differences in r 's for this sample. In earlier studies the authors did not break down the data by comparing children of different ages with their parents.

IQ's of true mothers (tested near the time the children were born) and their children who have grown up in foster homes (15). In this study the r 's for sixty-eight cases gradually rose from zero when the children were two years of age to .38 and .44 on the latest tests of the children, given when they averaged thirteen and one-half years.

The r 's for these two studies are, for several reasons, not directly comparable. However, the earlier study gives the general relationships to be expected when scores are available on both parents, and for a large population of children (age three years and older) growing up with their own parents. In the later study the comparisons are made repeatedly for the same population, the variable factor being the ages of the children at the time of testing. Although in this study the r 's are lower, they give evidence of increasing mother-child similarity in intellectual capacities as the children grow older. This is impressive in view of the fact that the children were separated from their true mothers in early infancy. Furthermore, correlations of the children's IQ's with educational status of foster parents remained close to zero (ranging at the thirteen and one-half year tests from .00 to .06).

In the absence of direct measures of intelligence, some other estimate of parental ability has often been used. We will concern ourselves here only with some of these studies which report correlations for children of different ages. In the Berkeley Growth Study (2, 6), and in the Guidance Study which has been reported by Honzik (10), several socio-economic factors have been correlated with the children's mental test scores at successive ages up to eight and ten years. In both studies the parent-child r 's are zero or even negative at first, but increase as the children grow older. Honzik found that by the time the Guidance Study children were eight years old, their mental test scores correlated .36 with mid-parent education, .41 with a socio-economic index, and .55 with a rating of parent's intelligence. By the time the Berkeley Growth Study children (2) were ten years old, their intelligence scores correlated around .55 with such factors as socio-economic ratings and education of parents.

PARENT-CHILD CORRELATION IN THE BERKELEY GROWTH STUDY

The Sample.—In the present report some parent-child comparisons are made on both mental and physical characteristics of the

TABLE I.—MEANS AND STANDARD DEVIATIONS OF COMPARABLE ADULT EDUCATIONAL AND PHYSICAL DATA FOR CHILDREN AND THEIR PARENTS

Heights (Inches)	N	Mean	SD
Mature* boys	22	71.63	2.38
Boys at 18 years	21	71.36	2.51
Parents of boys			
Fathers	22	70.57	2.46
Mothers	22	64.97	2.50
Mid-Parent†	22	70.37	2.11
All fathers	42	69.69	2.73
Girls at 18 or mature	21	65.29	2.57
Parents of girls			
Fathers	20	68.72	2.67
Mothers	20	63.92	2.33
Mid-Parent†	20	63.67	1.90
All mothers	42	64.47	2.48
Weights (Pounds)			
Mature boys	22	157.49	13.09
Boys at 18 Years	21	154.40	12.64
All fathers	41	161.71	19.30
Mature girls	24	129.87	19.45
Girls at 18 years	20	126.90	16.15
All mothers	42	132.76	18.74
Weight/Height‡ Indices†			
Boys at 21 years	17	218.8	17.5
Boys at 18 years	21	213.2	12.2
All fathers	41	233.0	25.1
Girls at 21 years	20	210.1	28.6
Girls at 18 years	20	212.6	26.6
All mothers	42	224.2	27.1
Education (Years of Schooling Completed)			
Boys	23	13.87	3.38
Girls	23	14.67	2.05
Fathers	43	13.66	3.82
Mothers	43	12.79	3.12
Mid-Parent	43	13.22	2.89

* Heights after closure of all epiphyses in hand and knee, often after eighteen years.

† Boys' mothers' heights multiplied by 1.08; girls' fathers' heights multiplied by 0.923.

‡ Indices computed from metric units.

children in the Berkeley Growth Study whose records are now complete through twenty-one years. The number of children tested at any given age varies from forty to sixty-one. Data are available for eighteen years or longer for forty-seven children from forty-three families. Table I gives the means and SD's for both parents and children in the characteristics to be considered here.

In heights and weights this group of parents is probably typical of California adults of similar age and racial antecedents. They are, for the most part, better educated, and have a generally higher socio-economic status, than a cross section of the Berkeley population with which they were compared at the outset of the Growth Study (6).

In comparison with their parents the children seem to be favored, both physically and mentally. They are, on the average, about an inch taller than their parents, but five pounds lighter. Consequently, they are slenderer in build, but this difference may be due to the fact that they are still younger than most of the parents were at the time the parental measurements were obtained.

Even though in a number of instances the children's educations are not yet completed, we find that they have had more education than their parents. This is true of the girls, in particular, but at least four of the boys are now in medical school or other graduate study. The mean IQ of these children at eighteen years on the Wechsler-Bellevue Scale was 122, SD 16 (3).

Mental.—As stated above, we do not have intelligence-test scores for the parents. Among the socio-economic variables studied, the parents' education was found to correlate most highly with the children's mental-test scores.

In Table II and Figure 1 these previously reported relations for the Berkeley Growth Study are reviewed and brought up to date. For graphical representation the product-moment r 's have been transformed to z 's.² In this figure, the z transformations are based on r 's between the midparent education (average of the years of schooling completed by mother and father) and the children's mental test sigma scores at every test age from birth through eighteen years. That is, we have compared a single evaluation of parental ability with a succession of different scores of their children's intelligence. It is evident that the parent-child correlation

² From the table on page 348 in McNemar (14).

TABLE II.—CORRELATIONS (AND Z SCORES) BETWEEN MID-PARENT EDUCATION AND CHILD'S MENTAL TEST SCORES AT EACH AGE TESTED

Age		Mid-Parent Education by	N	r	z
Mo.	1	California First-Year Scale	52	-.01	-.01
	2	California First-Year Scale	58	-.02	-.02
	3	California First-Year Scale	61	-.27*	-.28
	4	California First-Year Scale	58	-.22	-.22
	5	California First-Year Scale	58	-.29*	-.30
	6	California First-Year Scale	57	-.23	-.23
	7	California First-Year Scale	52	-.19	-.19
	8	California First-Year Scale	53	-.05	-.05
	9	California First-Year Scale	56	+.06	+.06
	10	California First-Year Scale	56	+.08	+.08
	11	California First-Year Scale	52	+.00	+.00
	12	California First-Year Scale	53	+.03	+.03
	13	California First-Year Scale	53	-.05	-.05
	14	California First-Year Scale	46	-.02	-.02
	15	California First-Year Scale	52	-.00	-.00
	18	California Preschool Scale	51	+.16	+.16
	21	California Preschool Scale	53	.29†	.30
	24	California Preschool Scale	48	.50†	.55
	27	California Preschool Scale	51	.41	.44
	30	California Preschool Scale	47	.44	.47
	36	California Preschool Scale	49	.47	.51
	42	California Preschool Scale	43	.35	.37
	48	California Preschool Scale	45	.50	.55
	54	California Preschool Scale	44	.59	.68
	60	California Preschool Scale	47	.58	.66
Yr.	6	1916 Stanford-Binet	48	.57	.65
	7	1916 Stanford-Binet	46	.50	.55
	8	1937 Stanford-Binet	47	.59	.68
	9	1937 Stanford-Binet	45	.57	.65
	10	1937 Stanford-Binet	47	.54	.60
	11	1937 Stanford-Binet	48	.61	.71
	12	1937 Stanford-Binet	46	.59	.68
	13	Terman-McNemar	41	.62	.72
	14	1937 Stanford-Binet	40	.60	.69
	15	Terman-McNemar	42	.56	.63

TABLE II.—*Continued*

Age	Mid-Parent Education by	N	<i>r</i>	<i>z</i>
16	Wechsler-Bellevue	44	.64	.76
17	1937 Stanford-Binet	43	.65	.78
18	Wechsler-Bellevue	40	.60	.69

* The negative *r*'s for months three and five are significant at the .05 level of confidence or better.

† The positive *r* at month twenty-one is at the .04 level of confidence. After this the *r*'s are all highly significant, usually above .001.

is at first zero and even becomes negative during the second quarter year (at month five the *r* of $-.29$ is significant at the .03 level of confidence). The *r*'s hover around zero from seven months through fifteen after which they rise rapidly to .50 at two years.

The positive *r*'s become statistically significant after eighteen months, the level of confidence being .01 or better from two years on.

When the sexes are treated separately, there appears to be little sex difference in the parent-child relationships in mental ability, in this group, although the girls' scores are oftener than the boys' significantly correlated with parental status. The *r*'s are given for ten ages in Table III.

Bayley and Jones, in an early report on these changing relationships (6), discussed the question of the probable contributions of hereditary and environmental factors. The evidence seemed to indicate that hereditary factors emerged gradually as the children matured, but the proportional contribution of environmental factors could not be determined. Conrad in 1931 (7) reviewed studies of kin resemblances in physique and intelligence. He concluded that when various relevant factors were controlled, the kin resemblances in physical and intellectual characteristics were about the same. He pointed out, however, that the mere similarity of correlation coefficients gives no proof of similar etiology of the resemblances.

Physical.—The question has arisen whether changes in parent-child similarities might occur during growth in other areas as well as in mental scores. Are there similar changes, for example, in physical characteristics which are relatively unaffected by environmental differences, given a sample of normally healthy children (1) within a normal environmental range? To test this, we have

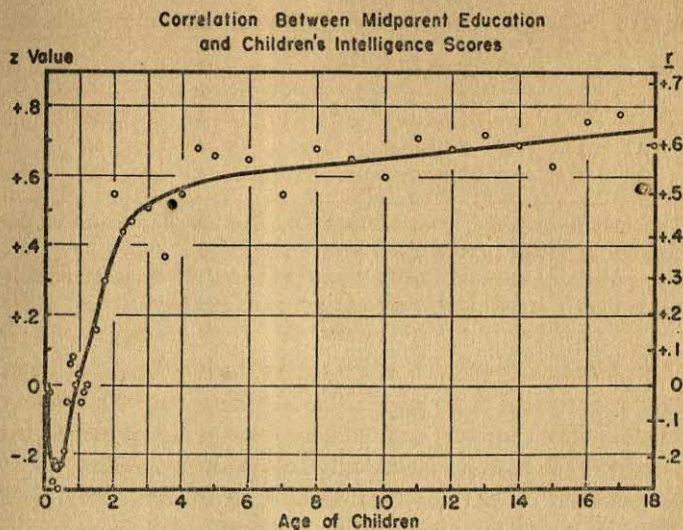


FIG. 1. Correlations between Children's Intelligence Sigma Scores and the Years of Schooling (Average of Mother and Father) of the Parents are Plotted by Age. Z-values are shown at the left, Correlation Coefficients at the right.

correlated some available parent-child physical data, i.e., heights and weights, for the Berkeley Growth Study.

The parents' heights and weights were obtained by the visiting nurse in an interview which occurred soon after the children were born, at a time when most of the parents were young adults in their twenties. The measurements were not made under laboratory conditions, but may be considered fairly good approximations. The children were measured nude, during their regular visits to the Institute, according to exact anthropometric techniques (5).

Correlations have been computed for three physical variables (height, weight, and an index of build, W/H^2), for the parents as adults with these same measurements of the children made at nineteen ages (and in some cases thirty ages). These r 's are given in Tables IV through VII. Because of physical sex differences, it is necessary to treat the sexes separately.

The parent-child correlations in height are given in Table IV. The r 's with sons tend to increase with age, although unevenly. The father-son r 's are significant after three years, but the mother-son r 's are not significant until thirteen years. The girls are significantly correlated with their mothers after the first few months, but not with their fathers, except possibly toward maturity.

TABLE III.—CORRELATIONS BETWEEN CHILDREN'S MENTAL TEST SCORES AND THEIR PARENTS' EDUCATION: SEXES SEPARATE

CA ①	N	Boys		N	Girls	
		Father	Mother		Father	Mother
		<i>r</i>	<i>r</i>		<i>r</i>	<i>r</i>
Mo. 6	31	-.29	.00	29	-.14	-.05
Yr. 1	28	.04	.24	28	-.08	.02
3	26	.39*	.49†	26	.26	.31
6	25	.34	.47*	26	.56†	.61†
8	24	.34	.40*	26	.64†	.59†
10	24	.44*	.38	26	.58†	.51†
14	21	.45*	.45*	21	.60†	.52*
15	21	.42	.42	21	.52*	.42
16	22	.48*	.40	22	.61†	.63†
18	21	.43*	.41	19	.60†	.63†

* Significant at the .05 level of confidence or better.

† Significant at the 0.1 level of confidence or better.

Midparent values for height have been computed by making an adjustment in the height of the parent of the sex opposite to that of the child (9). This is done by multiplying the heights of the boys' mothers by 1.08 and the heights of the girls' fathers by 0.923 and averaging the adjusted heights of the two parents. (In this sample the boys' mothers and fathers correlate with each other .31 in height and .04 in weight. For the girls' parents the corresponding *r*'s are: height, .10, weight .11). The correlations of the children's heights with the midparent values are shown in Table V and Figure 2. The boys show a continuously increasing correlation with their parents; the *r*'s become significant at about four years of age. The girls show an early rapid increase, becoming highly significant by five months, after which the *r*'s level off, then seem to pick up again after twelve years. However, from the data in Table IV, it would appear that these relationships with the midparent values are due primarily to similarities with the parents of the same sex.

The correlations for weight are shown in Table VI. The parent-son *r*'s are rather low, but there is a trend toward increasing relationship between the boys and their fathers, the *r*'s reaching the .05 level of confidence by fifteen years. The father-daughter *r*'s indicate some relation by five years, but they never reach statistical

TABLE IV.—CORRELATIONS (AND Z SCORES) BETWEEN HEIGHTS OF PARENTS AND THEIR CHILDREN

Age	Boys					Girls				
	N	Fathers		Mothers		N	Fathers		Mothers	
		r	z	r	z		r	z	r	z
Mo. 1	24	.30	.31	.05	.05	26	.36	.38	.30	.31
6	30	.25	.26	.27	.28	28	.42*	.45	.47†	.51
Yr. 1	28	.30	.31	.31	.32	28	.36	.38	.67†	.81
2	23	.28	.29	.21	.21	19	.34	.35	.56*	.63
3	25	.45*	.49	.25	.26	24	.38	.40	.67†	.81
4	22	.52*	.58	.25	.26	23	.35	.37	.66†	.79
5	23	.52*	.58	.30	.31	24	.24	.25	.70†	.87
6	24	.52†	.58	.35	.37	24	.26	.27	.75†	.97
7	24	.53†	.59	.35	.37	21	.33	.34	.76†	.996
8	23	.54†	.60	.43	.46	24	.26	.27	.74†	.95
9	23	.56†	.63	.43	.46	22	.30	.31	.71†	.89
10	23	.60†	.69	.37	.39	24	.24	.25	.72†	.91
11	24	.57†	.65	.41*	.44	21	.29	.30	.68†	.83
12	20	.53*	.59	.41	.44	22	.24	.25	.74†	.95
13	24	.45*	.49	.50*	.55	17	.56*	.63	.62†	.73
14	20	.58†	.66	.69†	.85	17	.49*	.54	.66†	.79
15	21	.42	.45	.71†	.89	20	.43	.46	.68†	.83
16	22	.47*	.51	.55†	.62	19	.42	.45	.66†	.79
17	20	.64†	.76	.60†	.69	20	.44	.47	.66†	.79
18	21	.52*	.58	.44	.47	19	.64*	.76	.52†	.58
21	17	.44	.47	.38	.40	20	.50*	.55	.67†	.81

* Significant at the .05 level of confidence or better.

† Significant at the .01 level of confidence or better.

significance. The mother-daughter correlations, however, show strongly increasing similarities in weight through sixteen years. They reach the .05 level of confidence with an r of .43 at two years and the .01 level by six years, when the r is .57, with r 's considerably higher at most of the later ages.

In our studies of growth we have found the W/H^2 index very useful as a measure of body build in children. It gives a measure of breadth, or volume, relative to stature. The parent-child correlations for this index are given in Table VII. The relative independence of the boys, especially in regard to weights, is reflected in

TABLE V.—CORRELATIONS BETWEEN MID-PARENT HEIGHTS AND HEIGHTS OF THEIR CHILDREN

Age	N	Boys		N	Girls	
		r	z		r	z
Mo. 1 ¹⁾	24	.21	.21	25	.46*	.50
2	32	.22	.22	27	.41*	.44
3	32	.24	.24	30	.42*	.45
4	31	.27	.28	29	.58†	.66
5	31	.23	.23	27	.56†	.63
6	32	.35*	.37	29	.61†	.71
7	27	.37	.39	24	.60†	.69
8	26	.36	.38	28	.48†	.52
9	28	.36*	.38	29	.67†	.81
10	29	.32	.33	27	.63†	.74
11	28	.36	.38	23	.51†	.56
12	29	.39*	.41	26	.67†	.81
18	23	.22	.22	23	.64†	.76
Yr. 2	27	.36	.38	23	.66†	.79
3	26	.44*	.47	23	.69†	.85
4	25	.52†	.58	23	.60†	.69
5	23	.50*	.55	24	.60†	.69
6	26	.56†	.63	24	.61†	.71
7	24	.54†	.60	22	.61†	.71
8	24	.58†	.66	24	.59†	.68
9	24	.60†	.69	22	.61†	.71
10	24	.60†	.69	24	.59†	.68
11	25	.60†	.69	21	.58†	.66
12	21	.59†	.68	22	.62†	.72
13	24	.62†	.72	17	.69†	.85
14	20	.82†	1.16	17	.69†	.85
15	21	.68†	.83	20	.67†	.81
16	22	.62†	.72	19	.67†	.81
17	20	.77†	1.02	20	.67†	.81
18	21	.58†	.66	17	.76†	1.00
21	17	.51*	.56	20	.69†	.85

* Significant at the .05 level of confidence or better.

† Significant at the .01 level of confidence or better.

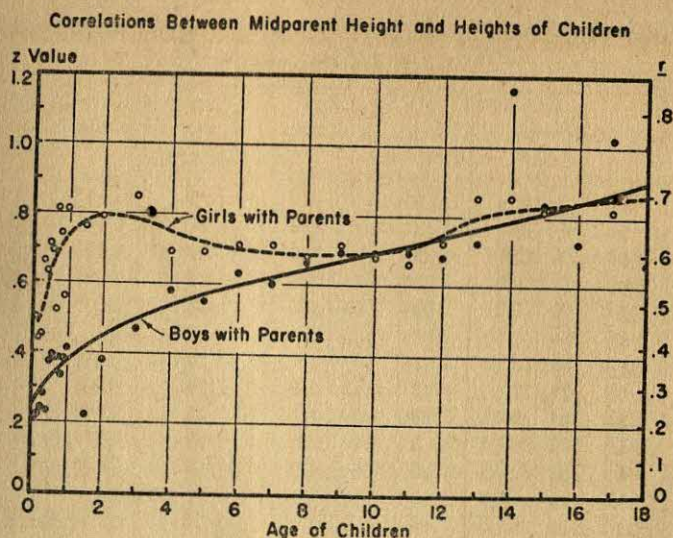


FIG. 2. The Solid Line and Dots Show the Z-values (and Correlations) between the Heights of the Boys and a Midparent Score (Boys' Mothers' Heights are Multiplied by 1.08 and Averaged with the Fathers' Heights). Broken Line and Open Dots Show these Values for the Girls and their Parents (Girls' Fathers' Heights are Multiplied by .923 before averaging with the Mothers' Heights).

these r 's: in the W/H^2 index the boys are unrelated to their parents at all ages. The girls' indices however, show increasing r 's with both parents. The father-daughter r 's approach significance at years eight, nine and ten, and again at sixteen and seventeen years. Mother-daughter r 's are usually above the .01 level of confidence at most of the later ages.

The sex differences in correlation are in some instances very striking. Even though the samples are small, statistical tests indicate that the differences may well be significant (Table VIII). In heights (Tables III and VIII) the boys are found to be more closely related to their fathers than to their mothers, at the .05 level of confidence or better, at years four, five, six, ten, and eleven; while the girls are more closely related (.05 level) to their mothers at all ages except thirteen and eighteen years, the difference being highly significant for years seven to twelve (.01 or better).

In general, also, the girls show more significant relation to their parents than do the boys. Their height correlations with midparent

TABLE VI.—CORRELATIONS (AND Z SCORES) BETWEEN WEIGHTS OF PARENTS AND THEIR CHILDREN

Age	Boys						Girls					
	N	Fathers		Mothers		N	Fathers		N	Mothers		z
		r	z	r	z		r	z		r	z	
Mo.	1	25	-.05	-.05	.35	.37	25	-.20	-.20	26	.41*	.44
	6	31	.13	.13	.23	.23	27	-.16	-.16	28	.23	.23
Yr.	1	28	.19	.19	.20	.20	27	.12	.12	28	.35	.37
	2	25	.26	.27	.02	.02	21	.01	.01	22	.43*	.46
	3	25	.33	.34	.07	.07	23	.24	.25	24	.47*	.51
	4	21	.23	.23	-.12	-.12	22	.19	.19	23	.41*	.44
	5	23	.34	.35	.09	.09	23	.37	.39	24	.47*	.51
	6	24	.32	.33	.23	.23	23	.30	.31	24	.57†	.65
	7	24	.23	.23	.20	.20	20	.45*	.49	21	.62†	.73
	8	23	.29	.30	.10	.10	22	.40	.42	23	.67†	.81
	9	23	.28	.29	.06	.06	20	.40	.42	21	.68†	.83
	10	23	.31	.32	-.01	-.01	23	.30	.31	24	.74†	.95
	11	24	.23	.23	.07	.07	20	.32	.33	21	.71†	.89
	12	20	.30	.31	.01	.01	21	.20	.20	22	.73†	.93
	13	24	.30	.31	.17	.17	16	.19	.19	17	.51†	.56
	14	20	.36	.38	.15	.15	16	.22	.22	17	.89†	1.42
	15	21	.46*	.50	.15	.15	19	.25	.26	20	.88†	1.38
	16	22	.45*	.49	-.09	-.09	18	.30	.31	19	.85†	1.26
	17	20	.49*	.54	.28	.29	19	.42	.45	20	.67†	.81
	18	21	.44	.47	.13	.13	18	.31	.32	19	.43	.46
	21	17	-.02	-.02	.10	.10	19	.16	.16	20	.81†	1.13

* Significant at the .05 level of confidence or better.

† Significant at the .01 level of confidence or better.

height (Tables IV and VIII) are significant at the .05 level or better for all but two of the fifteen *r*'s during the first three years: after that the boys' *r*'s are equally high. In weight (Tables V and VIII) girls are more significantly correlated with their mothers than are boys with their fathers, at thirteen of twenty ages, eight of these at the .01 level of confidence or better.

CHILDREN'S SELF-CORRELATIONS

It may be relevant here to point to the low relationship between these children's measurements in infancy and their own measurements as young adults. From Table IX, we see that mental test

TABLE VII.—CORRELATIONS (AND Z SCORES) BETWEEN W/H² INDEX OF PARENTS AND THEIR CHILDREN

Age	Boys						Girls					
	N	Fathers		Mothers		N	Fathers		N	Mothers		N
		r	z	r	z		r	z		r	z	
Mo. 1	24	-.15	-.15	.32	.33	25	-.15	-.15	26	.13	.13	
6	30	-.09	-.09	.23*	.23	27	-.31	-.32	28	.13	.13	
Yr. 1	28	-.12	-.12	.30*	.31	27	-.10	-.10	28	.07	.07	
2	24	.23	.23	-.11	-.11	19	-.04	-.04	19	.26	.27	
3	25	-.04	-.04	.16	.16	23	-.01	-.01	24	.14	.14	
4	21	-.27	-.28	.13	.13	22	.06	.06	23	.13	.13	
5	23	-.07	-.07	.26	.27	23	.19	.19	24	.14	.14	
6	24	-.24	-.25	.28	.29	23	.33	.34	24	.22	.22	
7	24	-.14	-.14	.31	.32	20	.39	.41	21	.39	.41	
8	23	-.18	-.18	.14	.14	22	.46*	.50	23	.44*	.47	
9	24	-.12	-.12	.04	.04	20	.45*	.49	21	.53*	.59	
10	23	-.04	-.04	-.16	-.16	23	.42*	.45	24	.64†	.76	
11	24	-.03	-.03	-.03	-.03	20	.31	.32	21	.75†	.97	
12	20	-.06	-.06	-.10	-.10	20	.37	.39	21	.65†	.78	
13	24	.09	.09	.06	.06	16	.40	.42	17	.35	.37	
14	19	-.08	-.08	.22	.22	16	.30	.31	17	.81†	1.13	
15	20	.17	.17	.16	.16	19	.31	.32	20	.78†	1.05	
16	22	-.09	-.09	.08	.08	18	.44	.47	19	.67†	.81	
17	20	.14	.14	.14	.14	19	.51*	.56	20	.41*	.44	
18	21	.14	.14	.03	.03	18	.47*	.51	19	.17	.17	
21	17	-.54	-.60	.25	.26	19	.43	.46	20	.69†	.85	

* Significant at the .05 level of confidence or better.

† Significant at the .01 level of confidence or better.

scores earned during the first year are unrelated to their eighteen-year test scores (3). As a matter of fact, the *r*'s of tests during the first nine months drop to zero or below when comparisons are made with scores at three years or older. However, the two-year and eighteen-years tests correlate .55, and starting at seven years all *r*'s with eighteen-year tests are above .80.

Growth in size is somewhat more stable, as may be seen in Table X. During the first year, the children's heights correlate in the neighborhood of .5 to .6 with their heights as adults. By the time they are three years old the heights of both sexes correlate above .83 with their own adult statures.

The self-correlations in weights, on the other hand, present a

TABLE VIII.—SIGNIFICANCE OF SEX DIFFERENCES IN PARENT-CHILD CORRELATIONS

		In Height			In Weight				
		Boys correlate higher with Fathers at this level of confidence*	Girls correlate higher with Mothers at this level of confidence*	Girls correlate higher with Midparent scores at this level*	Girls correlate higher with Mothers than Boys with Fathers at this level*				
Yr.	1	—	.01	Mo.	1	.03	Mo.	1	.01
	2	—	.04		2	.04	Yr.	2	.05
	3	—	.02		3	.04		3	—
	4	.03	.02		4	.02		4	.05
	5	.04	.005		5	.015		5	—
	6	.05	.02		6	.02		6	.03
	7	—	.005		7	.03		7	.01
	8	—	.003		8	—		8	.01
	9	—	.007		9	.01		9	.01
	10	.01	.003		10	.015		10	.005
	11	.05	.01		11	—		11	.004
	12	—	.003		12	.02		12	.006
	14	—	.05		18	.01		13	.05
	15	—	.03	Yr.	2	.02		14	.001
	16	—	.03		3	.02		15	.001
	17	—	.04					16	.001
								17	.04

* Not given if below the .05 level of confidence.

picture more like that of the parent-child relationships for weight. Although the self-correlations are pretty consistently higher than the same-sex parent-child r 's, the trends of correlation, with age, tend to be parallel.

Some correlations illustrating different degrees of consistency in build are shown for two indices in Table XI. The Stem Length/Height index is used to characterize the relative lengths of the legs and the body stem. The indices obtained on the Berkeley Growth Study cases at eighteen years are correlated with their indices at three earlier ages, three, twelve and thirty-six months. For both sexes the r 's are similar. They are lowest at one year, and in general show only a moderate relation to later status. The Weight/Height² index is correlated for the same age-intervals. This index of build shows no consistency in the boys between eighteen

TABLE IX.—CORRELATIONS BETWEEN EARLIER MENTAL TEST SCORES AND WECHSLER-BELLEVUE IQ AT 18 YEARS

Age	Test	N	r
Mo. 1	California First-Year Scale	36	.07
6	California First-Year Scale	36	-.12
Yr. 1	California First-Year Scale	35	.25
2	California Preschool Scale	36	.55
3	California Preschool Scale	36	.49
4	California Preschool Scale	34	.66
5	California Preschool Scale	36	.76
6	1916 Stanford-Binet	36	.77
7	1916 Stanford-Binet	36	.81
8	1937 Stanford-Binet, L	36	.85
9	1937 Stanford-Binet, L	34	.87
10	1937 Stanford-Binet, M	36	.86
11	1937 Stanford-Binet, L	35	.93
12	1937 Stanford-Binet, M	34	.89
13	Terman-McNemar, C	31	.93
14	Stanford-Binet, L	33	.89
15	Terman-McNemar, D	33	.88
16	Wechsler-Bellevue	36	.94
17	Stanford-Binet, M	36	.90

years and the first three years of life. The girls, on the other hand, appear to be very stable from one year.

The greater stability of the physical measurements and especially of heights, could be due to any or all of the following: (a) greater reliability of measurement, (b) greater stability during growth of the component parts in the values compared, and (c) smaller influence of environmental forces in producing individual differences, at least for this sample. Weight is very reliably measured, but it does vary with growth changes in body proportions, and with health and nutrition.

INDIVIDUAL TRENDS IN GROWTH

The changing status of a child, during his growth, may be illustrated by plotting a curve of his relative (or sigma) scores with age. Six such curves are shown in Figures 3 and 4. They include both mental scores and height scores, and on each child's chart the relative status of his own parents (compared with the parental

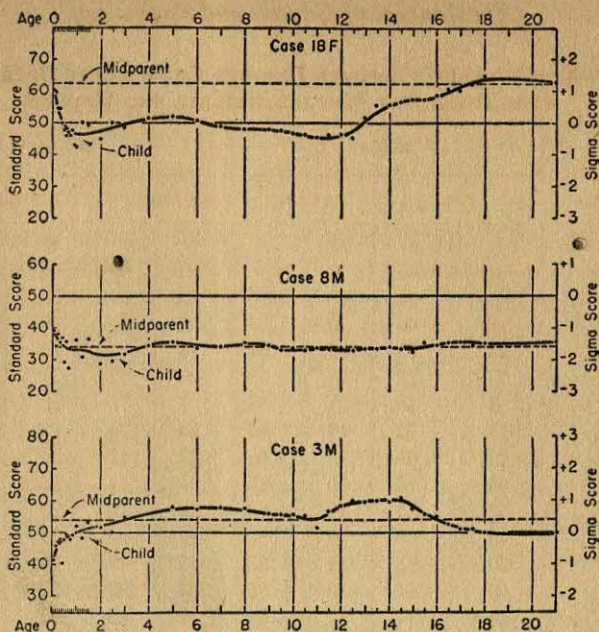
TABLE X.—CHILDREN'S EARLIER HEIGHTS AND WEIGHTS CORRELATED WITH THEIR OWN MATURE HEIGHTS AND WEIGHTS

Age	Height				Weight			
	Boys		Girls		Boys		Girls	
	N	r	N	r	N	r	N	r
Birth	—	—	—	—	23	.39	23	.51
Mo. 1	18	.64	20	.37	19	.19	20	.31
3	23	.53	23	.51	23	.18	23	.51
6	22	.72	21	.58	23	.32	21	.40
9	21	.59	23	.58	21	.12	23	.46
Yr. 1	21	.72	19	.65	23	.26	22	.51
2	21	.68	19	.68	21	.42	21	.62
3	21	.83	19	.85	23	.44	22	.59
4	21	.85	19	.82	20	.32	23	.63
5	21	.86	19	.84	22	.44	23	.68
6	21	.88	19	.85	22	.42	24	.79
7	21	.87	19	.85	23	.52	23	.81
8	21	.87	19	.84	22	.42	22	.87
9	21	.89	19	.83	23	.52	22	.86
10	21	.89	19	.80	22	.44	24	.87
11	21	.88	19	.64	23	.60	21	.81
12	21	.88	19	.77	20	.58	23	.78
13	21	.82	19	.88	22	.67	19	.53
14	21	.78	19	.73	19	.54	20	.86
15	21	.74	19	.98	20	.53	20	.92
16	21	.79	19	.997	22	.65	22	.95
17	21	.92	19	.999	20	.76	23	.94
18	21	.98	—	—	21	.89	—	—

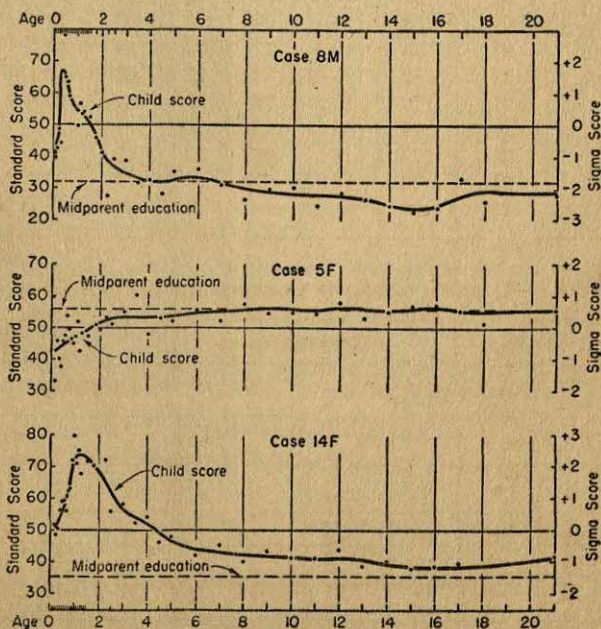
TABLE XI.—CORRELATIONS FOR CONSISTENCY OF TWO BODY BUILD INDICES: EIGHTEEN YEARS COMPARED WITH THREE EARLIER AGES

Age in Months	Stem Length/Height				Weight/Height ²			
	Boys		Girls		Boys		Girls	
	N	r	N	r	N	r	N	r
3	21	.47	17	.43	21	.10	18	.22
12	21	.08	19	.20	21	-.09	20	.81
36	21	.45	19	.61	21	.08	19	.71

STANDARD SCORES FOR HEIGHT



STANDARD SCORES FOR INTELLIGENCE



FIGS. 3 and 4. The Children's Sigma Scores relative to the Means for the Berkeley Growth Study for Successive Ages are shown by Dots and a Smoothed Curve. For Comparison the Parents' Sigma Scores relative to the Parents in this Group are indicated by a Straight Horizontal Line.

mean) is indicated by a broken horizontal line. As would be expected from the correlations, most (though not all) children's scores tend to settle at a level near that of their parents. But the individual curves bring out another fact, that some children reach that level much sooner than others.

The different patterns in physical growth can be shown to be due, in part, to differences in rates of maturing. For example, case 18F matured very slowly: her skeletal age was retarded for all ages after two years, usually lagging behind at least a full year. When she finally reached maturity she had attained a tall stature similar to that of her parents. Case 8M was rather consistently short, like his parents: his skeletal age was average until thirteen years, after which it lagged somewhat, apparently postponing his reaching an eventual status slightly taller than his parents. Case 3M, on the other hand, was accelerated at least a year at all ages after two years, and he tended to be relatively taller than his parents until he approached skeletal maturity. After sixteen years his scores fall below the parental line.

We do not understand clearly the nature of the maturing processes in mental functions. There could be inherent differences in rates, and also differential environmental factors operating to retard or enhance intellectual growth. But if we think of these children's scores at twenty-one years as measures of their mature intellectual status, it is clear that they follow varied paths in attaining that status. Case 8M had an early short period of rapid development, and after two years maintained a position that is close to both his parents' and his own mature level. Case 5F started out slowly, but caught up to her own adult and her parental level by three and one-half years. Case 14F, however, after a very precocious start settled more gradually to a position which never dropped quite as low as her parents' educational status, but stabilized at about eight years.

Examination of all the individual curves (in addition to those shown here) reveals a great variety of trends, with examples of (relatively) rapid or slow growth occurring at all ages, both in heights and in intelligence scores.

SUMMARY AND INTERPRETATION

Most of the physical comparisons show increasing parent-child similarities as the children grow older. However, there appear to be

some differences, both in the amounts of relationship and in the age-trends. Because the samples are small, it is possible that these differences are peculiar to the samples. If the samples are representative in these respects, it would appear that the within-sex relationships are stronger than those between parent and child of the opposite sex. Also, the mother-daughter relations seem to be closer than the father-son relationships.

The trends in mental relations differ in several ways from those in physical relations. The children themselves are more stable in the physical measurements. But the trends of increasing similarity to their parents seem more consistent in the mental scores. Although there are no sex differences in the parent-child r 's for the mental factors, there appear to be some marked sex differences in the physical factors. The analogy between the mental and physical relations is, obviously, not a close one. Furthermore, a variety of questions could be asked about genetic relations in physical growth which cannot be resolved here.

From these data it does, however, seem reasonable to assume the existence of a hereditary core of parent-child similarities in both mental and physical characters, even though such similarities may not be evidenced during the first year or so of the child's life. The similarities tend to increase as the children approach the mature status at which the characteristics were observed in their parents. Inherent differences in rates of development during infancy could account, at least in part, for the early lack of parent-child similarity.

Other factors operative in the reduced r 's at early ages are clearly observable in the phenomena of physical growth. These are the differences in the infants' body proportions from those of adults, and the differential growth rates of different parts of the body. For example, an infant's total height is made up in large part by his stem length, while the adult's height is determined to a much greater extent by the length of his legs. In the Berkeley Growth Study, the mean stem-length/height ratio drops from .65 at three months to .51 between twelve and fourteen years (signifying relatively longer legs), and then rises slightly to about .53 in the young adult. The transverse dimensions of the body, as contrasted with the longitudinal dimensions, grow more rapidly during the first year after birth and again during early adolescence.

Analogous age changes in organization could be expected in any

complex function during the process of its development (4). If we accept this position, then we may assume that differences in the mental organization of the infant and the adult could contribute to the changing parent-child relations in mental ability. Credence is lent to this hypothesis by the study of Skodak and Skeels (15) who report increasing mother-child r 's which cannot be attributed to environment. The effects of environmental influences should, consequently, be evaluated in the light of both age changes in organization (or relative maturity of the different parts or factors) and individual differences in over-all rates of maturing.

Although the sample used here is too small to allow for anything but tentative interpretations, it seems very likely that the changing parent-child relationships exhibited here are widespread. They deserve more extensive investigation in other populations and in other characteristics.

BIBLIOGRAPHY

- 1) N. Bayley. "Studies in the Development of Young Children." Berkeley: University of California Press, 1940, pp. 45.
- 2) ——— "Factors influencing the growth of intelligence in young children." *Thirty-Ninth Yearbook National Society for the Study of Education*, 1940, 39, 49-79.
- 3) ——— "Consistency and variability in the growth of intelligence from birth to eighteen years." *J. Genetic Psychology*, 1949, 75.
- 4) ——— "Development and maturation." *Theoretical Foundations of Psychology*, (H. Helson, Editor) New York: Van Nostrand Co. 1951, 145-199.
- 5) N. Bayley and F. C. Davis. "Growth changes in bodily size and proportions during the first three years; a developmental study of sixty-one children by repeated measurements." *Biometrika*, 1935, 27, 26-87.
- 6) N. Bayley and H. E. Jones. "Environmental correlates of mental and motor development: a cumulative study from infancy to six years." *Child Develop.*, 1937, 8, 329-341.
- 7) H. S. Conrad. "On kin resemblances in physique vs. intelligence." *J. Educ. Psychol.*, 1931, 22, 376-382.
- 8) H. S. Conrad and H. E. Jones. "A second study of familial resemblance in intelligence: environmental and genetic implications of parent-child and sibling correlations in the total sample." *Thirty-Ninth Yearbook of the National Society for the Study of Education, Part II*, 1940, 97-141.
- 9) H. Gray. "Prediction of adult stature." *Child Develop.*, 1948, 19, 167-175.
- 10) M. P. Honzik. "Age changes in relationship between certain environmental variables and children's intelligence." *Thirty-ninth Yearbook National Society for the Study of Education, Part II*, 1940, 185-205.
- 11) H. E. Jones. "A first study of parent-child resemblance in intelli-

gence." *Twenty-Seventh Yearbook, National Society for the Study of Education*, 1928, Part, 161-72.

12) H. E. Jones and H. S. Conrad. "The growth and decline of intelligence: a study of a homogeneous group between the ages of ten and sixty." *Genet. Psychol. Monog.*, 1933, 8, 233-297.

13) ——— "Parent-Child Resemblance in Mental Abilities as a Function of Age." Unpublished manuscript at the Institute of Child Welfare, University of California.

14) Q. McNemar. "*Psychological Statistics*." New York: John Wiley & Sons, 1949, pp. 364.

15) M. Skodak and H. M. Skeels. "A final follow-up study of one hundred adopted children." *J. Genet. Psychol.*, 1949, 75, 85-125.



TWO METHODS OF TEACHING ENGLISH COMPOSITION TO RETARDED COLLEGE FRESHMEN¹

RAY CHARLES MAIZE

Air University, Maxwell AF Base, Alabama

This study was established to provide a comparison between a 'language by experience' procedure and a 'grammar drill' method of teaching composition to retarded students in English A sections at Purdue University. Its purposes were: (1) to determine which method teaches the student more effective habits of language application as measured by objective testing and sample writing; (2) which method requires more extensive outside work and preparation on the part of the instructor; and (3) what relative factors, such as reading skill and mental levels, influence the student's ability to achieve success.

Sections of English A at Purdue University were used for the experiment. These slow and retarded students were separated from the other freshmen at Purdue by the scores they received on the American Council on Education Psychological Examination, the Purdue University Placement Test in English, and an original theme written during the entrance examination period. If their percentiles on the ACE Psychological Examination and the Purdue Placement Test in English averaged below 30 and their grade on the original theme (graded by the English faculty) was below 3 (passing), then they were assigned to English A, a non-credit, one-semester course which they must take before going on to English 1.

Recognizing the deficiencies of these students, the English faculty at Purdue had established English A as a remedial measure to bring the student to some degree of proficiency which would enable him to pass English 1, the required course for freshman credit in English. Taught according to the program established by the English department, English A consisted of workbook drill of grammar, punctuation, and spelling; and a weekly theme, corrected by the instructor. These procedures were supposed to be specifically de-

¹ This article is based on the writer's doctoral dissertation, Purdue University, 1952. Grateful acknowledgment is made to Professor Einar R. Ryden under whose direction this work was done.

signed to change the language habit patterns of slow, retarded freshmen.

To find out how well English A achieved the purpose for which it was established, permission was granted by the head of the English department to conduct an experiment in which the officially sanctioned method of instruction would be contrasted against a writing laboratory method. Accordingly, one group, the control, was organized and taught using the method established by the English department. The other group, the experimental, was taught using the writing laboratory method.

Inherent in this latter method would be the need to free the instructor from extensive theme reading, for it was recognized at the beginning that merely doubling or tripling the number of student manuscripts would provide no answer for the one great burden of teachers of English: how to get student writing activity without increasing the already too heavy load of the teacher. The second objective, then, of the experiment would be the need to prove that the teacher does not have to edit all student writing, that more effective results can be obtained if the teacher deliberately refrains from overdoing such supervision while at the same time encouraging voluminous efforts by the students.

The third objective of the experiment lay in the need to discover the special problems of the retarded students. What factors, in other words, had greatest relevance to their success in the course? Was it reading skill, mental level, vocabulary, attitude, or feeling of failure or success? Just as important, therefore, as the degree of control over his language habits which the experiment should reveal was the necessity to learn more about the individual who was to be taught.

STRUCTURE OF THE EXPERIMENT

From the freshmen students assigned to English A sections at Purdue during the year 1949-1950, one hundred and forty-nine were selected at random to become a part of this experiment to determine relative merits of two different approaches to the teaching of language skill on a college freshman level.

The method of selection for membership in any one of the groups has already been described. On the basis of these factors, students were registered by chance in the sections used for the experiment: seventy-four to the control group and seventy-five to the experi-

mental group. The control group was taught by the grammar workbook drill method; the experimental group was taught by the writing laboratory method.

During the semester of instruction, the following tests were administered: the Terman-McNemar Test of Mental Ability; the Henmon-Nelson Test of Mental Ability; the Iowa Silent Reading Test; the Rinsland-Beck Natural Test of English Usage, Tests I, II, and III; the Seashore Recognition Vocabulary Test; essays on an identical subject written before and after the period of instruction; the Horrall College Experience Scale; the Silance-Remmers Scale to Measure Attitude Toward Subject; and the Silance-Remmers Scale to Measure Attitude Toward Teacher.

To determine the amount of change and the significance of change before and after instruction, six units of measurement were established. These units were: (1) identical forms of the Seashore Recognition Vocabulary Test to determine change in total vocabulary; (2) equivalent forms A and B, Test I of the Rinsland-Beck Natural Test of English Usage, to measure change in ability to capitalize and punctuate; (3) equivalent forms A and B, Test II of the Rinsland-Beck to determine change in ability to work with diction, tense, pronoun reference, and verb forms; (4) equivalent forms A and B, Test III of the Rinsland-Beck to measure change in ability to improve sentence structure, organize logically, and write effectively; (5) combined scores for Test I, II, and III of the Rinsland-Beck to obtain composite language results; and (6) average of numerical theme grades awarded by two different English instructors after reading student manuscripts on an identical subject assigned before and after instruction.

Since one of the stated purposes of this study was to determine what relative factors have influence on the student's ability to achieve success, two different intelligence tests were used, in addition to the ACE Psychological Examination. The Iowa Silent Reading Test was also used to discover the relationship of scores in reading to intelligence scores and to other evaluation forms.

It was also considered important to find out what effect, if any, the personality and enthusiasm of the various instructors might have on the results. In addition, the attitude of depression or buoyancy, caused by success or failure, might also have an effect on the findings. For these reasons the Silance-Remmers Scale to Measure Attitude Toward Subject, the Scale to Measure Attitude

Toward Teacher, and the Horrall College Experience Scale were used.

During the period of instruction the control group wrote fourteen themes of approximately two hundred and fifty words in length. These themes were read and corrected by the instructor outside the classroom. The experimental group wrote forty themes of approximately two hundred and fifty words in length. The instructor did not read or correct any of this writing outside the regularly scheduled class hours; instead analysis and comment were provided in the writing laboratory by fellow students and the instructor.

ANALYSIS AND INTERPRETATION OF DATA

In this research the control and experimental groups were considered equal in ability at the beginning of the study. Data for indications of equality were obtained from the mean of the raw scores received by the students on the ACE Psychological Examination, the Purdue Placement Test in English, and the numerical theme grade. The t-ratios which were obtained for the difference between these means were not statistically significant according to t-tables.

MEASURE OF CHANGE

Table 1 indicates the results obtained from the six units of measurement which were established to determine the amount of change and the significance of change for the control and experimental groups.

As shown in Table 1, changes on the vocabulary test were favorable and significant at the one per cent level of confidence for both groups. A t-ratio for the difference between the changes, however, was not statistically significant. The difference between changes of the two groups could, therefore, be due to chance.

On Test I of the Rinsland-Beck the change was significant at the one per cent level of confidence for both groups. A t-ratio for the difference between changes, however, was significant at the five per cent level of confidence. The change for the experimental group, therefore, was significantly larger.

On Test II of the Rinsland-Beck the change for the control group was significant at the 5 per cent level of confidence, for the experimental group at the one per cent level of confidence. A t-ratio for the difference between changes was acceptable at the five per cent

TABLE I. Means and t-Ratios for Control and Experimental Groups on Measurements to Determine Change before and after Instruction

	Mean of Control		Mean of Experimental		t-ratio	
	Before	After	Before	After	Contr	Exp
Seashore Recognition Vocabulary Test	61.1	58.6	62.5	57.8	-3.6†	-5.9‡
Test I, Rinsland-Beck Natural Test of English Usage	55.4	59.0	55.0	62.8	5.1†	9.8‡
Test II, Rinsland-Beck Natural Test of English Usage	53.1	55.2	49.3	54.9	2.3†	5.6‡
Test III, Rinsland-Beck Natural Test of English Usage	29.5	28.3	24.9	33.2	-1.2	5.9‡
Combined Scores, Rinsland-Beck Natural Test of English Usage	138.1	142.4	129.2	150.9	3.1	12.8‡
Numerical Grades Given By Theme Readers	9.7	9.1	10.1	11.9	-1.2	6.0‡

† 1% level of confidence.

‡ 5% level of confidence.

level of confidence. The experimental group, therefore, achieved greater change than the control group.

On Test III of the Rinsland-Beck the control group showed no change that was statistically significant. The experimental group, on the other hand, showed change which was significant at the one per cent level of confidence. The t-ratio for the difference of change was also significant at the one per cent level. On this form, directly testing the student's ability to write, the evidence was all in favor of the experimental group.

On the Combined Scores of the Rinsland-Beck the total change for both groups was statistically significant at the one per cent level of confidence. The t-ratio for the difference of change, however, was significant at the one per cent level of confidence in favor of the experimental group.

On Numerical Grades Given by Theme Readers, the control group showed no significant change. The experimental group, on the other hand, did show positive gain in ability to write, significant at the one per cent level. The t-ratio for the difference of changes was also significant at the one per cent level.

CORRELATION OF FACTORS

One of the stated purposes of this study was to determine the relationship to one another of the many factors involved, in an effort to understand more clearly the nature of the teaching task and to obtain the necessary insight into those factors which seemed to contribute most to success or failure on the part of the student.

The following details are worthy of note:

a) The mean chronological age of the control group was 19.5 years; the mean age of the experimental group was 19.6 years.

b) The mean raw scores for both control and experimental groups on the Henmon-Nelson Test of Mental Ability were reported by the authors of the test at the 40th percentile, indicating that the students who form the population for this study were at least in the lower half of any population reported on by standardized tests.

c) When the mean raw scores on the Terman-McNemar Test of Mental Ability are compared with tables provided by the authors, the results show a mean IQ for both groups of 99 and a percentile ranking of 48.

d) The highest degree of correlation found between the Iowa Silent Reading rate and any other evaluation factor was that of .51 with the Q score on the ACE Psychological Examination. Otherwise the rate of reading seemed to bear little relationship with anything else.

e) The mean of the standard comprehension score as reported for the control and experimental groups on the Iowa Silent Reading Tests show a percentile ranking of 16 and a grade equivalent of 10.2, indicating a two to three year retardation in reading ability for students who are supposedly working on a 13th grade level.

EFFECT OF ATTITUDES

It was considered important to the results of this study to find out what effect, if any, the personality and enthusiasm of the various instructors might have. In addition, the attitudes of depression or buoyancy, caused by success or failure in the course, might also

have an effect upon the findings. It therefore became necessary to discover whether such moods differed between the control and experimental groups. On the Silance-Remmers Scale to Measure Attitude Toward Subject and Attitude Toward Teacher no difference could be found between attitude reactions of students in control and experimental groups. Similarly, no difference could be found on the Horrall College Experience Scale to measure depression or buoyancy. The conclusion was drawn that the lack of significant difference in measurement of attitude or mood indicated a lack of influence on the evaluation techniques used in this study.

SUMMARY AND CONCLUSIONS

As a result of the methods of instruction and the testing conducted before, during, and after the period of instruction, these conclusions can be drawn:

a) The control group showed statistically significant change measured before and after instruction on vocabulary, mechanics, and grammar. The experimental group, on the other hand, showed statistically significant change as measured before and after instruction on vocabulary, mechanics, grammar, controlled composition, and free composition.

b) Furthermore, the experimental group showed a statistically significant larger change on all measurements except vocabulary. All other measurements used in this study, therefore, showed clear evidence of superiority on the part of the experimental groups as the result of the methods of instruction.

c) All data revealed that students were of limited mental capacity and possessed an initial low level of ability in language use, vocabulary, and reading. In addition, changes for both low-level groups were slight, even after a semester of instruction.

d) No evidence could be found of the effect of student attitude or mood on the results of this study.

DIFFERENTIAL RESPONSES AMONG COLLEGE STUDENTS ON THE VOCABULARY SUBTEST OF THE WECHSLER INTELLIGENCE SCALE

CHALMERS L. STACEY and S. WILLIAM SPANIER

Syracuse University

Gerstein (1) in an article in the *Journal of Consulting Psychology* in 1949 suggests a relationship between the use of language and the underlying cognitive process involved. She proposes placing more emphasis on this relationship instead of merely determining whether or not the subject has given the correct definition of a word on a vocabulary test. That is, one should, in addition to determining whether the definition is correct or incorrect, attempt to differentiate the type of definition which is considered correct.

Reichard, Schneider and Rappaport (2) using the Goldstein-Weigl (3) color-sorting test as an instrument for investigating concept formation in children arrived at the conclusion that there exist three basic methods for forming concepts: (a) the concretistic, (b) the functional and (c) the conceptual.

Gerstein on the basis of the Reichard, Schneider and Rappaport study feels that the same three methods of concept formation which were exhibited in that study to solve a performance problem may also be used in classifying the definitional levels of responses to verbal stimuli. She suggests the following three classifications which parallel the aforementioned three: (a) the concretistic or descriptive method which is considered to be at the lowest level of verbalized definitions (the most primitive level), e.g., "an apple is red" or "a diamond is very hard and glitters, shines and sparkles"; (b) the functional or usage method, a more complex level and considered to be at the next higher level of verbalized definitions, e.g., "a coat is something you wear to keep warm" or "a microscope is used to make observations of minute objects"; and (c) the categorical or conceptual or abstract method, the most complex method and considered to be at the highest level of verbalized definitions, e.g., "wood is a fuel" or "a bicycle is a means of transportation".

STUDIES MADE ATTACKING THIS PROBLEM

In two separate studies, Stacey and Portnoy (4, 5) attempted to test the hypothesis put forth by Gerstein. Both studies used groups of individuals that were subnormal in intelligence as determined by the Wechsler Intelligence Scale. In their first study, when comparing the results of two groups of children, one classified as borderline and the other as mentally defective, they arrived at the following results, having used the Vocabulary Subtest of the Wechsler Intelligence Scale for Children for the verbal stimuli: (1) the borderline group (higher on the scale of intelligence) gave a greater proportion of descriptive type definitions than did the mental defective group—in fact, the two groups were significantly different at the one per cent level, (2) the mental defective group gave a greater proportion of the functional type of definition—this difference was also significant at the one per cent level, and (3) there was no significant difference between the two groups where the conceptual type of definition was concerned though the borderline group did show a slightly higher proportion of this type of definition.

In their second study, using three groups of adults, one classified as dull-normal, one as borderline and the last as moron, they found no significant differences in terms of the critical ratios between any two groups of the three in any of the three categories (descriptive, functional or conceptual). Nevertheless, the same general trend was evident. As in the first study, the higher level of intelligence gave a greater proportion of descriptive type definitions than did the next lower level of intelligence, and a lower proportion of functional type definitions than did the next lower level of intelligence, and a slightly higher proportion of conceptual type of definitions than did the next lower level of intelligence. Stacey and Portnoy, in this same study, then compared the two groups which comprised the two studies but omitting the dull-normal group of the second study, i.e., they compared the adult moron (of the second study) with the child moron (of the first study) and the adult borderline (of the second study) with the child borderline (of the first study) thereby holding intelligence constant and age as a variable. The following results were obtained: (a) the older age groups gave a greater proportion of descriptive and abstract types of definition, the difference being significant at the one per cent level, (b) the younger age groups gave a greater

proportion of functional definitions. Again the difference was significant at the one per cent level.

The results of these two studies appear directly to contradict Gerstein's hypothesis of the levels of methods of concept formation with the exception of the abstract or conceptual method. Gerstein's hypothesis is thus contradicted in terms of levels of intelligence and the Reichard, Schneider and Rappaport hypothesis (which includes the statement that children think at the concrete or descriptive level up to about age eight, and somewhere between the ages of eight and nine the functional level is reached, and at about the age of eleven the conceptual or abstract level is reached) is thus contradicted in terms of age.

In another study, Stacey and Markin (6) used the Similarities Subtest of the Wechsler Bellevue Intelligence Scale Form I with three groups of adults classified as (1) mental defective, (2) borderline and (3) dull-normal. Again as in the two previously mentioned studies by Stacey and Portnoy, the same trend may be noted. The greater proportions of descriptive responses were made by the higher levels of intelligence—the difference between the mental defectives and the dull-normal groups being significant at the one per cent level; the difference between the other groups, though not significant, showed the same trend. Also, the greater proportion of functional responses were made by the lower levels of intelligence—the difference between the mental defectives and dull-normal being significant at the one per cent level; the difference between the mental defectives and the borderline groups being significant at the five per cent level and the difference between the borderline and dull-normal groups, though not significant, again showed the same trend. The differences between the three groups at the abstract level were much too small to draw any conclusions.

Since all the results thus far have been gained through the use of individuals of subnormal intelligence as subjects, it is the purpose of this paper to further test Gerstein's hypothesis, but this time using as subjects one hundred eighty-eight college juniors and seniors.

METHOD

The vocabulary section of the Wechsler Bellevue Intelligence Scale was administered to one hundred eighty-eight college students. The responses given by the students to each of the forty-

five words in this section were recorded verbatim. The response sheet of each student was graded in accordance with the instructions cited in the Wechsler Bellevue Manual. Two separate groups of students were established by selection of the highest twenty-five per cent of the grades and the lowest twenty-five per cent of the grades (forty-seven students in each group). Each correct definition given by each student for each of the nouns (only nouns were used for analysis) listed on the response sheet was then broken down into one of the three categories: descriptive, functional or conceptual.

Each breakdown was then recorded in its proper column on a chart, and the total number of correct responses for each student also recorded on the chart. A grand total of correct responses was then calculated and recorded on the same chart. Separate totals were also calculated for each of the three categories and the proportions of each of the three separate categories to the grand total number of correct responses computed. This procedure was followed separately for the two groups—the upper twenty-five per cent and the lower twenty-five per cent. It might also be said that a third group was established from the middle twenty-five per cent of the grades and the same procedure followed with this group. Of course, any significant difference that might exist would come to light in the critical ratios computed between the high and low groups. The middle group should only serve to bridge the gap between these other two groups, and it is for this reason only that the third group was established. Critical ratios were then computed between the proportion of responses given by the three groups of students in each category, i.e., a critical ratio was computed for the proportion of descriptive responses in the high group as compared to the proportion of descriptive responses in the low group, and also the proportion of descriptive responses in the middle group was compared to the proportion of descriptive responses of both the high and low groups separately. The results of these comparisons are presented in Table I.

RESULTS

- 1) The proportion of descriptive responses made by the top and bottom groups differed significantly at the one per cent level. There was no significant difference between the proportions of descriptive responses given by the top and middle groups, but

TABLE I.—FREQUENCY AND PROPORTIONS OF LEVELS OF RESPONSE TO THE VOCABULARY SECTION OF THE WECHSLER-BELLEVUE INTELLIGENCE SCALE

Category	Top 25%		CR	Middle 25%		CR	Bottom 25%		CR	Top 25%	
	Freq.	Prop.		Freq.	Prop.		Freq.	Prop.		Freq.	Prop.
Abstract	1043	.74	4.05*	880	.67	2.11†	741	.63	6.11*	1043	.74
Functional	23	.02	6.66*	90	.07		78	.07	6.49*	23	.02
Descriptive	339	.24	1.21	347	.26	2.22†	347	.30	3.43*	339	.24
Total	1405			1317			1166			1405	

* Significant at the one percent level.

† Significant at the five percent level.

the trend was still evident—an increase in the proportion of descriptive responses given by the higher group to the next lower group. The proportion of descriptive responses between the middle and bottom groups differed significantly at the five per cent level.

2) The proportion of functional responses made by the top and bottom groups differed significantly at the one per cent level. The proportion of functional responses between the top and middle groups differed significantly at the one per cent level, but there was no difference at all in the proportions of functional responses between the middle and bottom groups.

3) The proportion of conceptual responses made by the top and bottom groups differed significantly at the one per cent level. The proportion of conceptual responses between the top and middle groups differed significantly at the one per cent level and the proportion of conceptual responses between the middle and bottom groups differed significantly at the five per cent level.

4) Each of the groups gave the greatest proportion of responses in the conceptual category, the next highest proportion of responses in the descriptive category and the lowest proportion in the functional category.

SUMMARY

The responses of the three groups of college students (groups arbitrarily established on the basis of scores on the Vocabulary Subtest of the Wechsler Bellevue Intelligence Scale) were analyzed as to the proportions of their vocabulary responses that fell into the descriptive, functional or conceptual categories.

According to the Gerstein hypothesis the descriptive type of definition is the most primitive, the functional type of definition is the next highest level, and the conceptual type of definition is the most complex. We should therefore expect that in any study of this kind we would find that the groups of higher intelligence levels would give (1) a greater proportion of conceptual type definition than would the groups of lower levels of intelligence, (2) a somewhat higher proportion of functional type definition than would the groups of lower levels of intelligence and (3) a lower proportion of the descriptive type of definition than would the groups of lower levels of intelligence. But in the three studies mentioned earlier in this paper, the results substantiated Gerstein's hypothesis in only one of the three categories—the conceptual. The other two categories—the functional and the descriptive—ran directly counter to the hypothesis set forth by Gerstein.

The results of this study are in agreement with the results of the previous studies on only two of the three categories—the conceptual and the functional. The discrepancy lies in the descriptive category. We therefore find that the results of this study substantiate Gerstein's hypothesis on two of the three categories—the conceptual and the descriptive. That is to say that: (1) the top group gave a higher proportion of conceptual responses than did the middle group and the middle group a higher proportion than did the bottom group, (2) the top group gave a lower proportion of descriptive responses than did the middle group and the middle group a lower proportion than did the bottom group; in contrast to Gerstein's hypothesis the top group gave a lower proportion of functional responses than did the middle group and there was no difference at all in the proportion of functional responses between the middle and the bottom group.

The discrepancy which exists between these results and those of the three previous studies—namely, the discrepancy in the descriptive category—may, of course, be due to the fact that in the previous studies subjects of subnormal intelligence were used whereas the sample in this study was composed of college students. Certainly a great deal more research is necessary before any absolute conclusions may be drawn or even assumed. However, the results of all four studies do leave in doubt that part of Gerstein's hypothesis which states that the functional type of definition is of a higher level than is the descriptive. But here again

further research is necessary before these results may be considered conclusive.

REFERENCES

1) R. A. Gerstein. "A suggested method for analyzing and extending the uses of Bellevue-Wechsler vocabulary responses." *Journal of Consulting Psychology*, 1949, 13, 366-370.

2) S. Reichard, M. Schneider, and D. Rappaport. The development of concept formation in children." *American Journal of Orthopsychiatry*, 1944, 156-161.

3) K. Goldstein and M. Scherer. "Abstract and concrete behavior: An experimental study with special tests." *Psychological Monograph*, 1941, No. 2.

4) C. L. Stacey and B. Portnoy. "A study of the differential responses on the vocabulary subtest of the Wechsler Intelligence Scale for Children." *Journal of Clinical Psychology*, 1950, 6, 401-403.

5) ———— "A study of the differential responses on the Vocabulary Subtest of the Wechsler Intelligence Scale." *Journal of Clinical Psychology*, 1951, VII, No. 2, 144-148.

6) C. L. Stacey and K. Markin. "A study of the differential responses among three groups of subnormals on the similarities subtest of the Wechsler Intelligence Scale." *American Journal of Mental Deficiency*, 56, No. 2, Oct. 1951.

ITEM VALIDITY AND RESPONSE CHANGE UNDER TWO DIFFERENT TESTING CONDITIONS¹

DOUGLAS G. SCHULTZ

Educational Testing Service

Princeton, N. J.

It has often been demonstrated that various responses to the usual self-description inventories and questionnaires can be elicited by altering the motivational pressures under which the instrument is administered. Bennett and Gordon (1), Ruch (5), and Schott (6) have all reported studies in which persons who took this type of test as part of the selection process for school or job gave considerably more favorable responses than persons who took the test under non-selection circumstances. Using the same questionnaire with which this report will be concerned, Myers (3) found a large number of statistically significant differences between the mean item responses of students accepted for college and those of college applicants. He also found that the applicants presented themselves generally as being more sober and industrious and better adjusted than the accepted students. Results obtained by Hollingworth (2), Simpson (8), and Van Wagenen (9) also support the point. When the answers given to inventories and questionnaires may have a decisive bearing on the attainment of some desired goal, it is apparent that many people either intentionally or unintentionally tend to respond in a manner which they think or feel will present the best opportunity for achieving that goal.

The study reported in this paper examined such shifts in response from the standpoint of the individual item. (Throughout the paper, responses to items and validities of items will be discussed.) The purpose of the research was to determine how changes in the item responses to an attitude-interest questionnaire administered under selection and non-selection testing conditions are related to the validities of the questionnaire items for an academic criterion. Specifically the two problems investigated could be

¹ This paper was read at the American Psychological Association meeting in Washington, D. C., on September 2, 1952.

phrased as follows:

1) Does the average response to an item change more (from non-selection to selection administrations) if the item is of high validity than if it is of low validity?

2) Do the items to which the average response is changed most also decrease most in validity?

HYPOTHESES

With regard to the first of these problems, it was anticipated that the correlation between mean item response change and external item validity would be zero, on the grounds that students seeking admission to college would exhibit no consistent insight into the way the items operate. It was felt that, in their general desire to give the most favorable impression, they would not differentiate between the more and less valid items and that therefore the mean change would be similar for items of all levels of validity.

Turning to the second issue, widespread response shifts from non-selection to selection groups were expected on all items and it was thought that these shifts might raise or lower item validities by chance. But it was hypothesized that there would be some tendency for the larger changes in item response in either direction to be associated with the greater reductions in item validities. Generally, then, change in mean response would be related to change in item validity only to a small extent.

SUBJECTS

Members of two successive freshman classes at a women's liberal arts college of an eastern state university were the subjects for the study. There were three hundred fifty-five women in the class of 1951 and three hundred forty-five in the class of 1952.

PROCEDURE

For several years the Educational Testing Service has been attempting to develop a questionnaire which would measure non-intellectual factors associated with academic achievement in college.² The study, supported by the College Entrance Examination

² The development and validation of this questionnaire over a three-year period are reported elsewhere in two papers (4, 7).

Board, has been directed primarily at evaluating characteristics of women entering college. A questionnaire of three hundred twenty-eight items was constructed and administered to the three hundred fifty-five members of the freshman class, the Class of 1951, as a group three weeks after their arrival at college. One hundred and forty-five of the same items were arbitrarily selected for reproduction in a booklet which was administered individually to applicants for the next year's freshman class, the Class of 1952, as a part of the normal application for admission; three hundred forty-five women were admitted to this class. Thus, members of the first class responded after they had been safely admitted to the college, whereas respondents in the second group were seeking admission to college. These are the postadmission and preadmission groups, respectively.

The one hundred and forty-five items in the questionnaire were of the usual inventory type. For example, one item asked: "Do you feel that the higher the grades a girl gets in college the more she will amount to after college?" The five possible responses were: (1) Yes, (2) Probably Yes, (3) Uncertain, (4) Probably No, and (5) No. In other items the respondent was to indicate how appealing certain topics of conversation might be to her. In some items she was to give the degree of importance to her of various reasons for attending college. Over all, the content covered the student's intellectual interests, motivation for attending college, and teacher relations and study habits.

Each of the items had four, five, or seven possible responses, numbered from 1 to 4, 1 to 5, or 1 to 7. Since this numerical order corresponded to a logical order of responses, a linear response continuum was assumed and the score of a person on an item was taken to be the number of the response which was checked. Thus, a respondent checking 'Probably No' in reply to the question cited above would receive a score of 4 on the item.

Using these scores, the average response for each of the one hundred and forty-five items in each of the two classes was computed. The Class of 1951 (postadmission) item mean subtracted from the Class of 1952 (preadmission) item mean was taken as an index of the change induced in the response to that item by the different conditions of administration. The difference for any item could be either positive or negative, depending only on whether

the average of the response numbers checked by the Class of 1952 was higher or lower on the arbitrary response scale than the average of the numbers checked by the Class of 1951. It should be noted that the same individuals are not being compared under two testing conditions here nor are two groups of matched pairs of students. Assuming that the two classes were selected in the same manner from the same general population, however, it is reasonable to compare the mean score of an item for one class with the mean score for the other class for the purposes of the study.

The criterion used for validating the items was the so-called 'achievement index', which was derived for each individual by taking the difference between her actual Freshman-Year College Grade-Point Average and her most probable Grade-Point Average predicted from the two scores of the College Board Scholastic Aptitude Test. (The multiple correlations of the Scholastic Aptitude Test verbal and mathematical sections with Freshman-Year Grade-Point Average were .51 in the Class of 1951 and .52 in the Class of 1952.) The product-moment correlation of each of the one hundred and forty-five items with the criterion was obtained. It was possible to calculate two such validity indices for each item, one in each class. It should be noted that since the items were of the common inventory type there was no 'right answer' in the sense of factual items and so there was no expected sign for any validity coefficient; either negative or positive values might be equally satisfactory and both were found in the present case.

Using the mean response differences and item validities described, two kinds of analyses were then carried out:

(a) Each item validity index in the Class of 1951 (non-selection) was plotted against the difference between the two class mean responses for that item; a similar plot was made with the Class of 1952 (selection) validity indices.

(b) The change in item validity from one class to the other was plotted against the mean item response difference. In considering the change in item validity a rise in the direction of the non-selection group's value (Class of 1951) was taken as a positive change or 'rise' in validity, a drop toward or beyond zero in the other direction as a negative change or 'drop'.

Since items of four, five, and seven choices were included, plots were also made substituting the critical ratio of the difference for

the mean item response difference. The item validity r 's were not transformed into z 's because all of them were low and so in the range where the transformation has little effect.

RESULTS

For the results, the two kinds of analyses will be considered separately. First, on the problem of item validity and item response change, the correlations computed from the scatterplots were as follows:³

In the Class of 1951 (Postadmission)

Item validity vs. mean item response change: $+.13$

Item validity vs. CR of mean item response change: $+.13$

In the Class of 1952 (Preadmission)

Item validity vs. mean item response change: $+.21$

Item validity vs. CR of mean item response change: $+.24$

The positive sign means that the largest item response changes in the direction indicated as desirable by the sign of the item validity coefficient are associated with the items of highest validity. Assuming that all the conditions underlying use of the common formula for the standard error of a zero correlation are fulfilled here, the N being one hundred and forty-five items, the two values of $+.13$ do not differ significantly from zero in the statistical sense, the $+.21$ almost reaches the .01 level of significance, and the $+.24$ does reach the .01 level. Although all four coefficients cited are positive, the pertinence of this agreement from the standpoint of statistical significance is somewhat reduced by the fact that the coefficients are by no means independent of one another.

As a matter of interest, the correlations using the critical ratio in an attempt to adjust for the variations in number of possible

³ The means and standard deviations for the four variables involved were:

	M	SD
Item validities (Class of 1951).....	-.02	.06
Item validities (Class of 1952).....	.00	.06
Differences between mean item responses.....	+0.05	0.35
CR's of differences between mean item responses.....	+0.57	3.99

A note should perhaps be added to clarify these item validities statistics. The item validities were distributed approximately symmetrically around zero, both positive and negative values being found as mentioned above, so the means of the item validities were close to zero as reported here. The most valid items, of course, would be those with the highest validity coefficients regardless of sign.

responses do not differ essentially from those using the straight-forward mean response difference. Also, the differences between comparable coefficients from the two classes are neither large nor statistically significant.

When the scatterplots for the four correlations were examined, the picture seemed to be a very common one with no suggestion of a curvilinear trend of any kind.

Turning to the second problem, the correlation between drop in item validity from the post- to preadmission class and change in mean item response was $+.01$. Substituting the critical ratio for the mean response difference resulted in a correlation of $-.02$. Thus, no association was found between decrease in item validity from one class to the other and change in mean item response. Here, too, examination of the scatterplots revealed only the common ellipse; a nonlinear relationship was not uncovered.

An interesting supplementary point arose out of the data involved in the two correlation coefficients just mentioned. The mean of the changes in item validity was $-.03$ (the standard deviation of the changes being $.06$). The negative sign attached to the mean reflects a general, slight drop or negative change (i.e., toward or beyond zero) in the item validities from the postadmission class to the preadmission class. How much of this drop was due to the unreliability of the item validity statistic and how much to the difference in situations is not known. The over-all distribution of item validities remained very much the same in the two groups, as one can see from the summary statistics in footnote 3 above.

CONCLUSIONS AND DISCUSSION

With regard to the two problems described earlier, the conclusions of the study are:

- 1) There was only a very slight tendency for the largest item response changes in the direction indicated as desirable by the sign of the validity coefficient to be associated with items of the highest validity.

- 2) No association was found between drop in item validity from one class to the other and change in mean item response.

While the results were derived from and apply directly only to the particular instrument involved, it seems reasonable that they might be found characteristic of other similar inventories, self-descri-

tion forms, and questionnaires in which the items are straightforward and not particularly subtle. Although students applying for college are apparently under some pressure intentionally or unintentionally to give what they consider to be desirable answers to these tests, such changes as occur are related only slightly, if at all, to the item validities. This suggests that the preponderance of student opinion as to which is the favorable end of the scale is correct to some extent but that there is great diversity of opinion in this respect about all the items. The suggestion from the second conclusion is that those items which are changed most from the postadmission to the preadmission situation are not any more adversely affected than are those which are changed least.

There is some comfort in these conclusions for questionnaire and inventory constructors. Even though the intent of this kind of item usually appears very obvious to the item writer, it does not follow that the student applying for college will necessarily be able to see through it. In his general effort to paint an attractive picture of himself the student seems rather unable, at least in the situation discussed here, to detect consistently the more valid items so that he may 'put his best foot forward' on these more important items especially. The emphases he injects into his questionnaire replies have little relationship to the validities of the items.

Another implication should be mentioned. The findings of the study raise the possibility that we need not necessarily discard a questionnaire or inventory for use in selection just because (1) the items in the instrument are not subtle and because (2) the normal item response for an applicant group is not the same as the normal response for an admitted group. Responses to the identical items may be quite different for the two groups, but many of the same items may be operative in differentiating among the members within each group. Where it is feasible, any instrument that appears to measure an important kind of behavior might well be worth trying even though the respondents are seeking a goal and might alter their responses accordingly. The empirical evidence will provide the evaluation required and proper use of the scores may still produce valuable information for selection purposes.

SUMMARY

An attitude-interest inventory was administered to two successive freshman classes at a women's college, in one case shortly after

the students had been admitted to college and in the other case as a part of the application procedure for admission. The average response change from the postadmission class to the preadmission class was computed for each item and plotted against the item validity for predicting freshman-grade performance relative to scholastic aptitude. The average item response change was also plotted against the change in item validity from one class to the other. There was only a very slight tendency for the largest item response changes in the direction indicated as desirable by the sign of the item validity coefficient to be associated with items of highest validity. No association was found between drop in item validity from one class to the other and change in mean item response.

REFERENCES

- 1) G. K. Bennett and H. P. Gordon. "Personality test scores and success in the field of nursing." *J. appl. Psychol.*, 1944, 28, 267-278.
- 2) H. L. Hollingworth. *The Psychology of Functional Neuroses*. New York: Appleton, 1920, Ch. VIII.
- 3) R. C. Myers. "A study of rationalization." *J. educ. Psychol.*, 1950, 41, 149-160.
- 4) R. C. Myers and D. G. Schultz. "Predicting academic achievement with a new attitude-interest questionnaire-I." *Educ. psychol. Measmt.*, 1950, 10, 654-663.
- 5) F. L. Ruch. "Ability of adults to fake desirable responses on two personality self-inventories and an attempt to develop a 'lie detector' key." *Psychol. Bull.*, 1945, 42, 539-540.
- 6) E. L. Schott. "Personality tests in clinical practice." *J. abnorm. soc. Psychol.*, 1937, 32, 236-239.
- 7) D. G. Schultz and B. F. Green, Jr. "Predicting academic achievement with a new attitude-interest questionnaire II." *Educ. psychol. Measmt.*, 1953, 13, 54-64.
- 8) R. M. Simpson. "Self-rating of prisoners compared with that of college students." *J. soc. Psychol.*, 1933, 4, 464-467.
- 9) N. B. Van Wagenen. *The Stability of Self-description Tests of Personality Adjustment*. New York: Author, 1935, pp. 46-47.

INFLUENCE OF SOCIAL CLASS UPON PERFORMANCE ON THE DRAW-A- MAN TEST

JOSEPH H. BRITTON

The Pennsylvania State College

Sociologists and cultural anthropologists are stressing the importance of the location of an individual in the social-class structure of his community as a basic determiner of a large proportion of the cultural and environmental experiences he may have (14, 15). By limiting the environmental experiences in socialization processes, a social-class group tends to restrict the individual's learning to a particular social pattern of behaving and believing, a pattern in many ways distinct from that of persons belonging to (or identifying with) other social classes (4, 12). The differences in cultural experience, together with an individual's cultural motivation, may be expected to influence the content of what a child will learn in school, how he will feel about his teacher, his peers, his school experience in general (3, 13, 17). These cultural differences may be expected to affect also the responses of children on intelligence tests (2). The problem of this investigation is to discover the effect of social class membership upon performance on certain standardized group tests of intelligence and upon performance on the Draw-a-Man Test (1).

Research on the relation of environmental-cultural factors and intelligence has been carried on by many investigators. Their work is reviewed by Kenneth Eells (5), as well as by Virgil Herrick (11). "The research is close to unanimous," writes Herrick, "in showing that there are significant differences in intelligence-test performance of children and youth from different socio-economic backgrounds, with children from the higher levels always securing the higher intelligence test scores." (p. 12).

A number of studies have shown the effect of environment upon the responses of children on the Draw-a-Man Test (8); among these is the research of Havighurst and his associates (9) concerning the performance of three hundred and twenty-five Indian children of the Sioux, Hopi, Zuni, Zia, Navaho, and Papago tribes in nine communities. They found that Indian children from these tribes

better than white children on the Draw-a-Man Test, that boys from the Pueblo groups did better than girls on the test, the evidence pointing strongly to the conclusion that environment affects the performance of children on the Draw-a-Man Test. To study the effect of social class environment on performance of white children on the Draw-a-Man Test is the problem of this investigation.

PROCEDURE

Subjects for the study include one hundred and two boys and one hundred and thirty girls whose social status in their midwestern community had been determined by using Warner's Index of Status Characteristics (16). Warner's index is based upon ratings, on seven-point scales, of four different status characteristics—occupation, source of income, type of house lived in, and dwelling area in the community. These four ratings are combined into a single index (ISC) by means of a weighting scheme; in this particular investigation, a part of a study of the University of Chicago (6), education of the parents was substituted for the source of income. The ISC has been shown to be closely related to social class placement as determined by actual social participation and social reputation in the community.

Standardized verbal and nonverbal group tests of intelligence (Henmon-Nelson, Otis Alpha Verbal, Otis Alpha Non-Verbal, Kuhlmann-Anderson) were administered to these nine-, ten-, and eleven-year-olds in their public and parochial school classrooms. The following year the children were asked to "draw the best man you can," and each of these drawings was scored twice on the Goodenough scale (7) by two assistants.¹

The comparisons between high and low socio-economic groups of boys and girls on the mean IQ's and their standard deviations were made for the five tests mentioned above. Tests of significance were made between status-sex groups on the Draw-a-Man Test. Then product-moment correlations were calculated between IQ's on the various tests and the ISC's.

¹ The aid of Miss Helen Coburn and Mr. David Twain, graduate assistants, Department of Psychology, The Pennsylvania State College, is gratefully acknowledged.

TABLE 1.—MEAN INTELLIGENCE QUOTIENTS AND STANDARD DEVIATIONS ON INTELLIGENCE TESTS FOR HIGH AND LOW SOCIO-ECONOMIC GROUPS OF 9-, 10-, 11-YEAR OLD BOYS AND GIRLS

Test	High SE Group*						Low SE Group†					
	Boy			Girls			Boys			Girls		
	N	Mean IQ	SD	N	Mean IQ	SD	N	Mean IQ	SD	N	Mean IQ	SD
Henmon-Nelson	52	115.0	14.4	67	114.1	16.3	41	94.2	13.4	55	98.2	15.2
Otis Alpha Verbal	54	106.9	9.4	66	102.8	10.4	43	96.3	9.9	56	93.8	10.0
Otis Alpha Non-Verbal	54	104.3	10.2	66	105.3	9.9	42	93.2	8.9	56	94.5	10.3
Kuhlmann-Anderson	53	107.5	8.7	69	108.6	10.1	41	93.0	10.5	58	96.8	11.3
Draw-a-Man‡	55	81.0	17.6	69	89.4	15.2	43	82.1	13.6	58	83.2	12.6

* ISC's 5-15 inclusive.

† ISC's 21-27 inclusive.

‡ Calculations were made to determine the statistical significance of differences between status-sex groups of mean IQ's on the Draw-a-Man test; all three differences involving the high status girls were significant.

RESULTS AND DISCUSSION

The results of the comparisons between high and low socio-economic groups are shown in Table 1. The high status group consists primarily of upper-middle and upper class boys and girls, and the low status group contains lower class children. It will be observed that both the high socio-economic group and the low socio-economic group were further divided into boys and girls. For both boys and girls of the high status group mean IQ's on the Henmon-Nelson, Otis Verbal, Otis Non-Verbal, and the Kuhlmann-Anderson are all above 100, the highest being the average IQ for the boys, 115.0. In contrast, the average IQ's for boys and girls of the low-socio-economic group all remain under 100, the highest in this case being the girls' mean Henmon-Nelson IQ 98.2. These results are very similar to those reported by Eells, of course, since our study population was a part of his. He found that the differences between his status groups, ranging from eight IQ points to twenty-three IQ points, were statistically significant. Our results are interpreted here to mean that clearly there are differences between the low and high socio-economic groups as to their performance on these widely-used intelligence tests. Moreover, as commonly observed, the differences favor the higher status group.

With regard to the mean Draw-a-Man IQ's for all four groups compared, they were definitely lower than the others; they ranged from 81.0 for the high status boys to 89.4 for the high status girls. Just why these average IQ's should be so low as compared to those of the other standardized tests is uncertain. There is, of course, the possibility of the pupils' not being motivated adequately enough to score well on the Goodenough; in gaining rapport in the classrooms the testers who aided us by administering the tests in the larger research project (6) endeavored to put the pupils at ease. The result may have been that they relaxed their efforts and felt that this was not a test but was designed simply to pass the time. Consequently, the results may be spuriously low. It should be pointed out, however, that in some instances the variability of the groups on the Draw-a-Man test, as shown by the standard deviations, is larger than on other tests.

Tests of the significance of differences between mean Draw-a-Man IQ's revealed that all those involving the high status girls were beyond chance expectations, those girls' average IQ being significantly higher than others. In Havighurst's investigation (9) involving the Indian children, a contrasting sample of Midwest white children was obtained; within that group there was a slight superiority of girls over boys on the Draw-a-Man Test, though the difference, they say, may have been due to chance.

Further average IQ's on the Draw-a-Man were calculated and are shown in Table 2. The difference between the mean IQ of the total high status group and that of the total low status group was not statistically significant; the difference between the mean IQ's of all boys and all girls was significant at the one per cent level.

Table 3 shows the results of correlation analysis of the IQ scores on the various tests with the Index of Status Characteristics. The

TABLE 2.—MEAN DRAW-A-MAN IQ'S AND SD'S SELECTED GROUPS

Group	N	Mean IQ	SD
High S-E group	124	85.6	16.5
Low S-E group	101	82.7	13.1
All boys*	98	81.5	15.5
All girls*	127	86.5	14.5

* The difference between mean IQ's of all boys and all girls was statistically significant.

TABLE 3.—PRODUCT-MOMENT CORRELATION COEFFICIENTS OF ISC'S WITH IQ'S ON VARIOUS TESTS

Test	N	r and PE _r
Henmon-Nelson	222	.51 ± .04
Otis Alpha Verbal	226	.42 ± .04
Otis Alpha Non-Verbal	223	.50 ± .04
Kuhlmann-Anderson	228	.53 ± .03
Goodenough Draw-a-Man	232	.11 ± .04

four coefficients involving the Henmon-Nelson, Otis Verbal, Otis Non-Verbal, and the Kuhlmann-Anderson are all moderate, r 's ranging from .42 to .53. Each r is statistically significant, being larger than four times its probable error. The higher the social status, as measured by the ISC, the higher is the intelligence, as measured by these IQ tests. This tendency is in conformity with other investigations concerning the relationship between intelligence and socio-economic background.

The coefficient showing the correlation between the response on the Goodenough Draw-a-Man Test and the ISC is .11 with its probable error indicating that this small degree of relationship is unreliable. In contrast, then, with the previous tests, this shows little or no relationship between the Draw-a-Man IQ and socio-economic status. Performance on the Goodenough scale does not appear to be influenced by the children's membership in a particular social class group. This finding may be substantiated by data in a study of ten-year-old white children by Havighurst and Janke (10), whose data do not lend themselves to rigid comparison with ours. In the midwestern community they studied, some differences were revealed between the status groups on the Goodenough scale. In the small samples they compared on this particular test, the mean IQ for lower-lower class children was 91; for lower-middle class children the mean IQ was 107, the difference between these socio-economic groups being 16.

One of the assumptions which test constructors utilize is that all subjects taking the test must be equally familiar or equally unfamiliar with the test materials. In fact, Goodenough's selection of a man as the object of this test was on the basis that all children (presumably all children in American cultures) have had approximately equal opportunities to observe the details of the human

figure on which scoring was standardized, so that resulting variations could be said to be due to discrepancies among children in their capacity to make use of such opportunities, rather than to their not having had such opportunities.

A legitimate hypothesis might be to the effect, however, that on the Draw-a-Man Test children of higher socio-economic status would perform somewhat better than lower class children. This we felt might be true because the higher status group would probably have had more experiences with drawing materials and with pictures of people in books and magazines; their relationship with school teachers would probably have been more positive, and the degree of their motivation to do well on the test higher than that of the lower class children. Moreover, the attention most middle and upper class parents give to details of clothing, personal neatness, etc., might be expected to influence their children's awareness of specific details which in turn might increase appreciably their scores on the Goodenough.

This difference in performance by social class groups, we might further hypothesize, would be less than with the other tests analyzed because of the almost complete non-verbal nature of the Goodenough and its administration, as compared with tests using verbal and picture-type items, which have been shown to favor often-times the upper status groups.

It would appear that on the basis of our data, social class environment has not affected to any significant extent the responses on the Draw-a-Man Test; but the differences which are apparent between the correlation coefficients of the Henmon-Nelson, Otis Verbal, Otis Non-Verbal, the Kuhlmann-Anderson, and the Draw-a-Man do appear to substantiate the latter hypothesis on the relative influence of social class. It is believed that the nature of the test materials themselves is the largest contributing factor. A significant difference between performance of low status groups and higher status groups on the Draw-a-Man was not revealed.

SUMMARY AND CONCLUSIONS

In this paper we have summarized findings of a study of the performance of some elementary school children on certain widely used intelligence tests. Having obtained their families' socio-economic status ratings in their midwestern community, by means of Warner's Index of Status Characteristics, we were able to compare

their IQ scores by social class. Further, having administered and scored Goodenough Draw-a-Man Tests on these children, we were able to show the relationship which existed between performance on the Draw-a-Man Test and social status. This association proved to be negligible; the relationships between social status and IQ's on the Henmon-Nelson, Otis Alpha Verbal, Otis Alpha Non-Verbal, and the Kuhlmann-Anderson were moderate but statistically significant.

The widespread use of intelligence tests as selectors of human abilities—as instruments to help educational counselors in guiding students and to aid industrial and military leaders in selecting personnel—is well-known. We submit that the optimum utilization of human resources, actual and potential, in all socio-economic groups is an obligation. The search for measures of intelligence which resolve social-class differentials must, therefore, be continued; questions of motivational factors which influence the performance of children on tests should receive concerted attention along with research on the content of tests themselves.

REFERENCES

- 1) J. H. Britton. "Influence of social class upon performance on the Draw-a-Man Test." *Amer. Psychologist*, 1952, 7, 304. (Abstract)
- 2) A. Davis. *Social-class Influences upon Learning*. Cambridge: Harvard University Press, 1949.
- 3) ———. "Socialization and adolescent personality." In N. B. Henry (Ed.). *Forty-Third Yearbook of the National Society for the Study of Education*. Part 1, *Adolescence*. Chicago: Department of Education, University of Chicago, 1944.
- 4) A. Davis and R. J. Havighurst. *Father of the Man: How Your Child Gets His Personality*. Boston: Houghton Mifflin, 1947.
- 5) K. W. Eells. *Social-status Factors in Intelligence-test Items*. Unpublished doctor's dissertation, University of Chicago, 1948.
- 6) K. Eells, A. Davis, R. J. Havighurst, V. E. Herrick, and R. Tyler. *Intelligence and Cultural Differences: A Study of Cultural Learning and Problem-solving*. Chicago: University of Chicago Press, 1951.
- 7) Florence L. Goodenough. *Measurement of Intelligence by Drawings*. Yonkers, N. Y.: World Book Company, 1926.
- 8) Florence L. Goodenough and D. B. Harris. "Studies in the psychology of children's drawings: II 1928-1949." *Psychol. Bull.*, 1950, 47, 369-433.
- 9) R. J. Havighurst, Minna Gunther, and Inez Pratt. "Environment and the Draw-a-Man Test: the performance of Indian children." *J. abnorm. soc. Psychol.*, 1946, 41, 50-63.
- 10) R. J. Havighurst and Leota L. Janke. "Relations between ability

and social status in a midwestern community. I. Ten-year-old children." *J. educ. Psychol.*, 1944, 35, 357-368.

11) V. E. Herrick. "What is already known about the relation of the IQ to cultural background?" In K. Eells, et al. *Intelligence and Cultural Differences: A Study of Cultural Learning and Problem-solving*. Chicago: University of Chicago Press, 1951.

12) W. L. Warner and Associates. *Democracy in Jonesville: A Study of Quality and Inequality*. New York: Harper & Brothers, 1949.

13) W. L. Warner, R. J. Havighurst, and M. L. Loeb. *Who Shall Be Educated?—the Challenge of Unequal Opportunities*. New York: Harper & Brothers, 1944.

14) W. L. Warner and P. S. Lunt. *The Social Life of a Modern Community*. Vol. 1. *Yankee City Series*. New Haven: Yale University Press, 1941.

15) ———. *The Status System of a Modern Community*. Vol. 2. *Yankee City Series*. New Haven: Yale University Press, 1942.

16) W. L. Warner, Marchia Meeker and K. Eells. *Social Class in America: A Manual of Procedure for Measurement of Social Status*. Chicago: Science Research Associates, 1949.

17) *Progressive Educ.* 1950, 63, No. 4 (whole issue).

CLASSROOM ENVIRONMENT AND PUPIL WELFARE

NEWELL C. KEPHART

Purdue University

and

WILLIAM FLOYD

West Lafayette (Indiana) Public Schools

Recently public school systems have given considerable attention to the physical environment of the school classroom. Traditional color schemes with their preponderance of dull shades of gray and buff have given way to brighter and more cheerful decorative motifs. Woodwork, floors and furniture have changed from traditional dark oak to lighter shades of sand or bleached finishes. Rigid desks in straight rows have been replaced by movable desk and chair combinations which can be adjusted to the physical requirements of the pupil. Control of both natural and artificial light is distributing the quantity of available illumination more evenly and increasing its usefulness for each pupil in the classroom.

In the Morton Elementary School, West Lafayette, Indiana, a series of research studies was undertaken to determine the effects of these new classroom designs upon the welfare of pupils who spent a school year in this new environment. The room design used followed that of the 'coördinated classroom' recommended by Harmon.

Two classrooms were refinished in accordance with these more modern developments. Three major changes were undertaken. First, the distribution of light was rearranged so that the quantity and quality of light was distributed equally, as nearly as possible, over all portions of the room. Care was taken to see that at any seat in the room adequate light was available for the critical recognition of visual detail and the full and free performance of visual tasks, bearing in mind that such visual tasks are not confined to a single plane but are three dimensional in character. An optimum distribution of light was obtained by installing semi-trans-

parent baffles at an angle to the upper half of the windows in such a manner that two thirds of the light entering through the window was reflected onto the ceiling and one third was transmitted in the normal manner. The baffles were set at such an angle that the reflected light from the ceiling was directed to the normally dark sides of the room furthest from the windows. The ceiling and an eighteen-inch strip at the top of the walls were painted white to aid in this distribution of light. Natural light was supplemented by artificial light also designed to be distributed equally over all parts of the room. For this purpose florescent fixtures were installed in reflectors so designed that the light was evenly distributed.

The second consideration in the refinishing was the amount of contrast between the visual task presented to the child and its surrounding. It is desirable that this contrast be reduced markedly from that which occurs in the normal classroom and preferably be not greater than three to one. Since most schoolroom tasks involve working with white paper, the desks in the experimental room were finished in a light wood finish which reduced markedly the contrast between the paper on which the child was working and the surrounding desk tops. For the same reason the finish on the floor was light. Dark woodwork was replaced by light finish and pastel shades were used in the painting of the walls. Black chalkboards were replaced by yellow green chalkboards to further reduce undesirable contrast.

The third change involved the furniture in the room. It is desirable that such furniture permit freedom of movement to the child and be adjustable both to the physical size of the child and to the task at hand. The American Seating Company ten-twenty desk was chosen and was used to replace the previous furniture. These desks were not placed in the customary straight rows but were arranged in slightly curved rows so that as nearly as possible a fifty degree angle between the plane of the windows and the child's work was maintained. Such an arrangement minimizes the undesirable shadows which fall on a pupil's desk.

Two additional rooms identical to the experimental rooms, as to location on the side of the building, number of windows, etc., were chosen as control rooms. For each maintenance operation which was performed in the experimental classroom, a similar operation was performed in the control classroom. When the experimental classrooms were painted, the control classrooms were

also painted but the traditional color scheme was followed. When new furniture was provided in the experimental classrooms; new furniture, but of a more traditional design, was provided in the control classrooms. Similarly, with all other necessary alterations in the experimental classrooms, a similar amount of work, material, and money was put into the control classroom. As a result the control rooms were just as new, just as clean, and just as completely refinished as were the experimental rooms. The only difference was that the scheme followed in refinishing the control rooms was traditional whereas the scheme followed in refinishing the experimental rooms was that of the coördinated classroom.

One experimental room and one control room were designed to be occupied by fourth-grade classes. The second experimental room and the second control room were designed to be occupied by fifth-grade classes. At the beginning of the school year, all students enrolling in the fourth grade were assigned at random to either an experimental or a control room. Similarly all students enrolling in the fifth grade were assigned at random to either an experimental or a control room. By this procedure, two randomly selected groups were obtained at each grade level, one assigned to the experimental room and one assigned to a control room. This random selection resulted in two groups in each grade which did not differ significantly in educational achievement level or IQ.

No attempt was made in this series of experiments to evaluate all of the possible effects of the experimental room environment upon school children. Similarly no attempt was made to investigate all of the hypotheses which have been advanced in support of the coördinated classroom. Rather, three general areas of investigation were chosen which, it was thought, were important to the welfare of school children. The three areas selected were school achievement, posture, and frequency of nervous habits. An experimental design was developed in each of these three areas which it was felt would throw light upon the question of differences between the experimental and control groups. The results of these three experiments are reported here.

EDUCATIONAL ACHIEVEMENT

In investigating the variable of school achievement, scores on the Stanford Achievement tests, intermediate examination and the Henmon-Nelson test of mental ability were used. Only the two

fifth-grade classes could be used for this experiment since the tests administered at the fourth-grade level were not comparable. These results are therefore based on twenty-three children in the fifth-grade experimental class and twenty-four children in the fifth-grade control classroom.

For each child the accomplishment quotient was computed. This quotient evaluates educational achievement in terms of both chronological age and mental ability and, therefore, represents achievement corrected for the mental ability level of the pupil. These quotients were then averaged for each of the two classes separately. For the experimental group, the mean accomplishment quotient was 92.1. For the control group, the mean accomplishment quotient was 85.4. The difference between the two groups was 6.7 in favor of the experimental classroom. This difference was found to be statistically significant with a *t*-value of 3.48.

From these results it is seen that the educational achievement as measured by the accomplishment quotient was greater in the class which spent the school year in the experimental environment than in the case of the class which spent the school year in the control environment. The difference between these two classes was significant at less than the one per cent level of confidence.

POSTURE

In investigating postural differences between the two groups of children, only one aspect of total posture, head-tilt, was considered. A head-tilt is defined as a tendency on the part of the child to hold his head tilted noticeably to either one side or the other when he is standing erect and performing a visual task which is located at eye level and twenty feet in front of him.

To measure this variable, a photographic technique was used. A 35 mm. camera was mounted on a tripod and placed directly in front of the child but below his eye level. Across the room at a distance of twenty feet a Snellen visual acuity test chart was mounted on the wall at the child's eye level. The child was brought into the photographic room and was shown a circle eighteen inches in diameter drawn upon the floor with chalk. The Snellen chart across the room was pointed out to him, and he was instructed to stand in the circle and read the letters on the chart. No further instruction was given as it was desired not to influence the

dividual's posture by the instructions. When the child had read down the Snellen chart to the point where the letters became difficult for him to recognize, a full length front view photograph of him was taken.

When these photographs had been developed and printed, judges were requested to determine whether or not a head-tilt was present for each child. The instructions to the judges were to identify the midline of the body by observing the sternal notch and a point midway between the knees. They were to observe the midline of the face by determining a point midway between the eyes and the point of the chin. If the midline of the body and the midline of the face coincided, the judge was to indicate that a head-tilt was not present. If the two lines formed an angle, he was to indicate that a head-tilt was present.

From an observation of the photographs, it was apparent that the measurements represented a bimodal curve. If a head-tilt was present, it was present to a marked degree and it was clearly recognizable. There were extremely few cases in which the head tilted so slightly as to leave doubt in the minds of the judges. There appeared to be very few cases of slight or doubtful head-tilt. For this reason the measurement of these photographs by the method of judgment or rating and the establishment of a bi-category variable appeared to be justified and was followed in computing further results.

The reliability of the rating method for determining posture described above was investigated statistically. In this investigation, two problems arose. The first problem was the reliability with which different raters would evaluate the same set of photographs. To test this aspect of reliability, two raters were asked to judge the same set of photographs. Their results were then compared using the tetrachoric r as a measure of agreement. This procedure resulted in an r of .97.

The second reliability problem involves the degree to which the postural characteristics revealed by the photographs would remain constant from day to day. To test this problem, a group of children were photographed on one day using the procedure described above. On the following day the same children were photographed again using the same identical procedure. A rater was asked to judge both sets of photographs according to the procedure outlined above. The result of his judgment on these two sets of

data were then compared, using the tetrachoric r as the measure of agreement. This procedure resulted in a coefficient of .81, which indicates that the postural characteristics revealed by this photographing technique are relatively constant from day to day and consequently indicate an habitual posture which persists and tends to recur whenever the child is performing a task similar to that represented by the present experimental situation.

When the reliability of the photographic technique and the scoring method had been established, the number of children with head-tilt was determined for all groups in both grades and in both environments. There were thus a total of forty-nine children who had spent one full year in the control environment. Of these control children, forty-nine per cent displayed a head-tilt when measurements were taken in May. There were also forty-nine children who had spent one school year in the experimental environment. Of these experimental children, only twenty-seven per cent showed a head-tilt when measurements were taken in May. There was thus a difference of twenty-two per cent between the groups. This difference was statistically significant, the t value being 2.18. It would appear that the experimental environment has a less detrimental effect upon the habitual posture of the child as indicated by the presence of head tilts.

FREQUENCY OF EXTRANEIOUS MOVEMENTS

Subjective observation of the two types of classroom environment involved in these studies had suggested that children in the control classrooms might show a higher frequency of extraneous movements or nervous habits. In order to investigate this possibility, the technique developed by Olsen was used. This technique consisted of an observation of a timed sample of behavior. Olsen had constructed a list of the most common nervous habits displayed by young children. The technique consisted of observing a child for a predetermined period of time and noting for each habit in the list, the number of times which it occurred during the observation period.

Using Olsen's list of nervous habits, similar observations were made in the four classrooms involved in the present study. In September, and again in May, an observer went into the classroom. With Olsen's list before him, he observed the behavior of the first child in the room for a two minute period. Whenever one

the habits on Olsen's list was observed, a tally mark was made on the scoring sheet. At the end of the two minutes, the observer transferred his attention to the next child whom he observed for a similar two-minute period. These observations were continued until all of the children in the room had been observed. The total number of tally marks on each score sheet was determined and this figure was considered the raw score on the nervous habits inventory for that child. The score thus represented the number of times that activities listed by Olsen occurred during a two-minute observational period. Each of the four classrooms was treated in the same manner.

In order to determine the reliability of this technique, two observers were sent into each of two rooms and made observations according to the procedure outlined above simultaneously on each child in the room. The results of correlating the ratings by each of the two raters was a reliability coefficient of .93.

In order to determine any difference between the two types of classrooms, the results were treated separately for the control rooms and the experimental rooms. In the control rooms, the average frequency of nervous habits in the September observations were computed. A similar average was computed for the May observations. The difference between these two averages was noted. This difference was found to be $-.63$, indicating that the average frequency of nervous habits was slightly higher in May than in September for this group. This difference however, was not statistically significant as the t value was only .95.

Similar averages were obtained for the experimental groups and the difference computed. The difference in these experimental groups was found to be 1.52 , indicating that the frequency of nervous habits was less in May than in October. These experimental group differences were found to be statistically significant, having a t value of 2.71 .

It is thus seen that the children who spent a school year in the control type classroom showed no change in the number of nervous habits which they displayed in a two-minute time period. On the other hand, the children who spent this school year in the experimental classroom environment showed a decrease in the number of nervous habits which they displayed in a similar two-minute time period.

It appeared that the experimental environment was having some

effect upon the number of extraneous movements or nervous habits which the children displayed. If it can be argued that these nervous habits are an indication of the level of tension under which the individual is working, then it can be similarly argued that this tension is less in the experimental type classroom than in the control type classroom.

SUMMARY AND CONCLUSION

This study attempted to evaluate the effect upon fourth- and fifth-grade school children of a school year spent in an experimental classroom. The experimental classroom used was that recommended by Harmon and known as the coördinated classroom. A group of children who spent a school year in this type of classroom was compared with a similar group who spent the same school year in a traditional type classroom environment. Three areas of development were studied: school achievement, posture, and frequency of extraneous movements or nervous habits. It was found that the group who spent the school year in the experimental environment was superior to the group who spent the school year in the control environment in school achievement as measured by the accomplishment quotient, in the posture of the head in relation to the body (a smaller percentage of head-tilts), and in extraneous behavior patterns (a lesser frequency of extraneous movements or nervous habits). In the areas investigated in this study, it is concluded that the newer type classroom environment represented in this case by the coördinated classroom can have a beneficial effect upon school children who spend a considerable portion of their time in this environment during the school year.

BOOK REVIEWS

ROGER G. BARKER, BEATRICE A. WRIGHT, LEE MEYERSON, AND MOLLIE R. GONICK. *Adjustment to Physical Handicap and Illness: A Survey of the Social Psychology of Physique and Disability*. New York: Social Science Research Council, (Bulletin 55th Revised), 1953, pp. 440.

The social psychology of physique and disability is a problem that has obtained in ever increasing extent the interest of both research people and workers in the rehabilitation field. Bulletin 55, the original volume of the Social Science Research Council which surveyed this topic has influenced the direction as well as the amount of interest and work on the subject. This appeared in 1946. Enough changes have taken place since, in the opinion of the investigators, to warrant a revised edition. The pattern of organization, selection and sequence of topic considerations is the same as the first. The findings are reported in six chapters with the same captions, and a bibliography follows. There are changes, and the changes are largely in terms of what the investigators consider to be changing trends calling for a reappraisal to serve as a guide for still further research and still more realistic and needed applications.

The outstanding change in this revised edition is in the shift away from emphasis on statistical facts and correlations to a larger emphasis on the use of human documents. In the revised edition there is a definite improvement in literary style. The revision is more than merely a repair job. It is that, but it is more. The changes made are sometimes just in rephrasing of a word, reconstruction of a phrase, and at times a paragraph. The most marked changes are in the first two chapters: The introductory chapter on methodological problems of somatopsychological research is described in some detail with a much more substantial perspective for comprehensive orientation for understanding all the problems than is done in the original volume. Also Chapter 2 concerned with normal variation in physique is presented in a fuller fashion. To a small extent, such differences as appear between the original edition and the present revised edition are similar in nature to those found in the two volumes of Middletown by the Lynds: *Middletown* full of statistics

and with the omission of human factors and *Middletown in Transition* beginning the book with the Ball family and describing the people behind the scenes. The changes here are not that drastic and marked, but they are in the same direction of increased significance of research and applications rather than a mere listing and tabulation. The chief changes are in a broader comprehensive orientation in defining the field both in increased clarity of concepts used and an increased searching for theories of soma-psyche relations, so there's a more frequent searching for and explanation of likely explanatory hypotheses. Also revealed in the more recent studies is improved technical adequacy of some of the research. These changes for the better are stressed and influence the revised edition. But they also do one other thing—they bring to the focus of attention some of the urgent needs that still remain in the field. Everything is considered related to the problem they are concerned with that, in their way of defining it, is "Concerned with those variations in physique that affect the psychological situation of a person by influencing the effectiveness of his body as a tool for actions or by serving as a stimulus to himself or others." The list of physical variations which can have significance as listed by them are eighteen in number and include size, race, age, sex, cosmetic defect, muscular strength, motor ability, speech defects, auditory impairments, visual impairments, tuberculosis, heart disease, diabetes mellitus, rheumatism, leprosy, cancer, orthopedic disabilities, acute illness. Also included in their consideration are some psychoanalytic contributions such as Schilders in consideration of the body image in relationship to the self-concept. And the investigators call attention to one practical phase of the problem; namely, the urgent need for investigating the effectiveness of forms of coping with the problems of disabled workers.

All in all, the investigators made a competent survey of the literature in their first effort, and the people who are in the field, either in the field of rehabilitating and educating the disabled or social scientists interested in the followup of studies of social psychology of physique and disability, will find this volume an even more satisfactory frame of reference for orientation in the field than was the first volume.

Psychological Service Center

H. MELTZER

MARJORIE B. SMILEY. *Intergroup Education and the American College*. New York: Bureau of Publications, Teachers College, Columbia University, 1952, pp. 212, \$4.25.

The foundation for this latest volume of the Teachers College Studies in Education is provided by a detailed and intelligent analysis of the intergroup education programs reported by fifty-six colleges and universities in response to questionnaires submitted by the author. Each chapter presents the philosophy of a certain aspect of intergroup education, a summary of policies and practices of the fifty-six institutions in regard to this aspect, and recommendations for planning a curriculum to deal with this aspect. The organization should be convenient for individual college instructors or committees charged with the task of working out a curriculum concerned with desirable attitudes, values and skills in intergroup relations. To such committees and teachers the book can be recommended. The emphasis on curriculum development rather than methodology makes it of limited value to most workers in the social sciences.

Neither the author's thoroughness nor her command of the subject and related areas will be challenged by many readers of this volume. Profuse quotations and citations of authority plus a completely-thought-through concept of intergroup education leave no doubt of the writer's familiarity with the subject. However, this style of writing is probably also the major weakness of the volume. Perhaps the author felt the necessity to reinforce her arguments with the statements of authorities, because she is aware that some college teachers might not see the need for intergroup education. Whatever the reason, the style makes for low readability from the standpoints of both interest and difficulty, and may discourage some readers.

University of Oklahoma

WILLIAM ELLER

✓ BERNARD H. HALL, MARY GANGEMI, V. L. NORRIS, VIVIENNE HUTCHENS VAIL AND GORDON SAWATSKY. *Psychiatric Aide Education*. New York: Grune and Stratton, 1952, pp. 168.

Psychiatric Aide Education is a report of a seven-year experiment of a training program carried out by the Menninger Foundation in cooperation with the Topeka State Hospital. A grant from the

Rockefeller Foundation supplied the necessary funds. The school opened officially on October 3, 1949, and closed on June 30, 1952. A total of eighty-five students were matriculated; sixty of them graduated. The story of this project includes a report concerning the field of psychiatric aide—its growth and its needs; recruitment and selection of students; an evaluation of the curriculum; a story concerning its graduates; and an evaluative chapter on recent developments in psychiatric aide education. Also included is a fair sized appendix, which contains a bibliography and forms used or relevant in considering the problem of psychiatric aide education. The forms include application for admission, course outlines, and course evaluations.

All in all, it's an interesting story written by the staff members. Dr. Karl Menninger in his preface describes the significance of this project as well as his reaction to the obstacles experienced because of prejudice and blind spots in unexpected places. This story should be of interest to hospital administrators and those concerned with the problem of psychiatric aide education.

Psychological Service Center

H. MELTZER

EDWIN FORBES TAIT. *Textbook of Refraction*. Philadelphia: W. B. Saunders Company, 1951, pp. 418.

This book on refraction by a professor of ophthalmology at Temple University is primarily intended for a student group preparing for the practice in the field of eye work. Accordingly, the author has sensibly enough selected to write the book in straightforward, at times seemingly dogmatic form, using references that are pertinent and readily available. Also included by him are books and papers suggested as further reading on the more original material that can be used by more ambitious students or the more curious and inquiring readers.

The text as written presupposes a background of fair knowledge of the anatomy and physiology of the eyes and the nervous system as well as an acquaintance with the principal methods of geometric and physiological optics. The early chapters contain discussions of the basic material necessary for the study of refractive error and ocular motor anomalies. They also contain a short, elementary presentation of the more important concepts which are discussed in greater detail later in the volume. There is a chapter on vision

that continues with a consideration of preliminary concepts for understanding problems in the field and then chapters are outlined that are of anatomical structure or of functions related to the field. Illustrative of topics included are: ocular reflexes; anomalies and variations of ocular accommodation; myopia; astigmia; presbyopia and aphakia; binocular fusion, and principles and methods in heterophoria.

In all, this is a carefully prepared volume that should serve well to do what it proposes to; namely, serve students in the field as well as neurologists and psychologists who are interested in having an authoritative volume on ocular refraction.

Psychological Service Center

H. MELTZER

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 45

February, 1954

Number 2

EDUCATIONAL THERAPY

BRUCE McCARTY

Waco, Texas

This is a report on the Experimental Educational Therapy program which was conducted at the Austin State Hospital from March to the end of August of 1951. It was cosponsored by the Hospital and the Hogg Foundation for Mental Health.

RATIONALE

Mental institutions have been thought of historically as places where unfortunates were kept away from civilization for their mutual protection. Not until comparatively recently has it been thought possible to restore mental patients to a healthy active rôle in society, and with the dawn of this concept, the name of the institution was changed from 'asylum' to 'hospital'. Sad to relate, the practice has not kept pace with the theory. In many instances the professed intention of relieving mental illness and restoring the individual has taken a back seat to administrative convenience which is often suffused with the file-and-forget philosophy. More often than not, various therapies which are in use are merely peripheral measures designed more to occupy time than to ameliorate aberrances.

It seems likely that even though the etiology of mental disorders is primarily emotional, the patient's preoccupation with these personality problems interferes with his intellectual accomplishments and this in turn causes him to fall behind in his own rate of development—in what Havighurst calls 'developmental tasks'. Though these are frequently thought of as being problems of a strictly academic nature, such is only part of the story; sundry items such as the acquisitions of graceful interpersonal relationships, advancement in current activities (whether educational, occupational or

personal in nature) and just generally keeping up with the-~~here-~~and-the-now are no less developmental tasks which must be mastered in their proper sequence to form stepping stones toward future life manipulation. The befoulment of this orderly sequence of developmental progress in turn causes an emotional problem of ~~his~~ ^{his} own. As the individual sees himself being left behind by the onrushing world his urge for inferiority-avoidance reaches near panic proportions and the vicious cycle thus created is not materially relieved by therapies designed to waste yet more time.

Clearly, if these two facets of the problem are simultaneously operative they must be attacked simultaneously in order to render therapy effective. There must be (1) the traditional psycho-environmental attack on the emotional problem: environmental in order to release the patient from the tension causing rigidity of his old milieu, and psychotherapeutic to rebuild an ego substantial enough to withstand the stresses and strains of life.

Concurrently, there must be provided (2) the opportunity to bring expectations and accomplishments into line. It is hard to see how there could be any successful and permanent rebuilding of the ego with only the materials which were present from the incipient stages of the morbid condition. These were evidently inadequate as a bulwark once, and there is little reason to suppose them adequate as time and tide go on apace. The addition of tools for living is essential, whether academic, professional and/or avocational; and in its widest sense we call this increment Educational Therapy.

CLASS PLAN

The idea of Educational Therapy for the younger patients of the Texas State Hospital was primarily a pilot study for assessing the effects of education as a means of therapy for the students, and also the effect of these abnormal students on education.

The curriculum was academic- rather than vocational-centered. Of course this does not mean that it was built exclusively around the three 'R's'. A typing program, for instance, radio course and art classes were pointed in the direction of specific trades, but the emphasis was on the traditional subjects.

The school was set up in the former office of the beauty shop. This room was particularly well suited for the purpose in that it was light, modern-looking, well ventilated and out of the heavy

traffic flow. The walls were a combination of stained wood and plaster, the latter part of which was used as a movie screen. One of the two long tables contained our book collection and three archaic typewriters (one of which was in such condition that it served only in the capacity of a bookend). The other table was work space.

Books on arithmetic, art, astronomy, chemistry, civics, English, geography, geometry, health, history, physics, radio, spelling and trigonometry were provided by the Special Education section of the Austin Public School System. Other equipment included water colors, crayons, white and colored chalk, construction paper, tablets, typing paper, molding clay, pencils, finger paints, magazines and scrapbook making material.

LIMITATIONS

All too often a mental institution is a catchall for everything society deems undesirable enough to segregate from itself. In order to make a hospital an instrument of rehabilitation some degree of discrimination must be maintained. If it is necessary, because of inadequacies in the legal system, for the hospital to cater to the subnormals as well as the abnormals, the hospital must make a frank differentiation between the programs to be followed in treatment of these two categories. The purpose of Educational Therapy programs is to restore the patient at least to his former level of efficiency, and, if possible, to increase his level of functioning; but it must be recognized that no educational technique invented has ever replaced a missing gray cell. By trying to make one program flexible enough to include every form of deviation the net result will be reduced to zero.

This is not intended to relegate the subnormals to the wastebasket. It merely segregates them from the group under discussion. The subnormals form a group which is custodial in nature, and their inclusion in a dynamic program for readmission to normal society will lead only to a blind alley.

The idea of having the school open to voluntary attendance by any and all the hospital patients was advanced but seemed impractical on at least six counts. (1) The nature of the students makes it imperative that the classes be small so that individual attention can be given to the students. On several occasions this has been borne out substantially when it was found that two students could not be given simultaneous instruction on par

assignments. This is not strange in that, perhaps, those who became students in this school were those who were unable to function adequately in their society, thus it is small wonder that they find group participation difficult—even on this limited scale.

(2) Another difficulty which is closely related to the above is the definite amount of work which an assignment can contain for each individual. If any assignment exceeds this amount it is rejected as being too difficult, and the student feels 'put upon'. If the assignment contains less than the optimum it is within easy reach of the student which creates for the teacher the problem of keeping everyone occupied. With a large class the teacher cannot get around to all before they have finished such assignments.

(3) These students are more insistent than the normal students about having their work checked and graded immediately. They lack courtesy and tact and will interrupt any activity in progress to secure their most recent desire. A large class of this type would soon deteriorate into a condition of bedlam.

(4) Another practical disadvantage to having 'open' class is the disquieting effect (to say the least) created when the teacher walks into a room filled with students about whom he knows nothing except that they must be abnormal or they wouldn't be there. Not knowing their educational achievements or their idiosyncrasies the teacher is at a loss to know how and where to start and the direction to take.

If the teacher starts a paranoid too low his pride is injured and rapport is extremely difficult to establish subsequently. The same amount of coercion-cajoling that can be used successfully with a schizoid meets with disaster if used on a hostile organic student.

A positive program of admission to the school which was worked out and adhered to circumvented many of these unfortunate situations. Opportunity for the teacher to brief himself on each case before the student was admitted to the classroom saved confusion and embarrassment in all directions.

(5) Equipment, books especially, are taken to the wards at will because it is felt that this permission gives the students a sense of belonging and being accepted. The free flow of students, however, makes this arrangement obviously inadvisable.

(6) Several of the students are interested in making the class a meeting place for themselves and their *objets d' amour*, and little

or no academic ambition enters the picture. They easily become a disrupting influence for, if they are put to work, they take the limited facilities from those who have genuine academic aspirations, and if they are not put to work, they undermine the entire tenor of the class with their erotic activities.

Thus it was felt to be in the best interests to limit the students in the class, and to this effect a screening criterion was set up. It might be noted that as things turned out this standard was frequently honored in the breach. Four categories of patients were set aside as being inadvisable: (1) those of advanced age. These were sufficiently removed from school by virtue of their years that it was felt their inclusion in school would tend to be a regressing factor rather than a normalizing factor. In the case of the younger patients, even though their academic progress might be slight by comparison to outside standards, it was felt that for them to engage in an activity which resembled, at least to some degree, their concept of normal outside work would be in some measure therapeutic.

(2) Those currently engaged in a course of shock treatment were eliminated also. The advisability of this was subsequently underlined by the exception. One girl who was engaged in shock therapy disrupted classes thoroughly; it was impossible to arrest her attention even for short intervals, she could remember none of the names of the classmates (even though her avowed true love was among them) and in general she macerated academic progress.

(3) Those with organic involvement (brain damage), especially of the progressive types, were placed on the nonpreferred list because it was thought that their inclusion with others of greater capacity would be beyond the point of diminishing returns with respect to their administration. There were three, however, who were admitted to the class: one trauma and two lobotomies. The trauma case did little of academic interest and was much more concerned with sex relations than arithmetic. One lobotomy made great strides toward regaining proficiency in geometry, but the other was on shock treatment also and the summative effect was that she was considerably beyond the pale of advisability for our class.

(4) The psychopathic personalities were frowned upon mostly because of their anti-social activities. When all is said and done,

it is still necessary to maintain some sort of order in the classroom. To maximize the effectiveness of the program, it was thought advisable to minimize the number of psychopaths.

AIMS AND PURPOSES

Education as therapy upholds both the principles of modern education and the tenets of good mental health; for both are based on the same concept—capitalization on individual differences—but neither science escapes modification by the union. Our purpose is not purely education. If it were this experimental project would be unnecessary; any typical classroom situation would suffice. Our students were frankly abnormal—beyond the limits of tolerance in normal classrooms; so it seems advisable to define the aims and purposes which guided the experimental class.

- 1) To provide basic academic skills for the student to use to facilitate his future adjustment.

For anyone to succeed in this day as even a somewhat autonomous creature, he must have some control over his environment. It is held to be almost patent that one who has attained a chronological development far in excess of one's academic progress is at a distinct disadvantage in one's dealings with inanimate objects and especially with interpersonal relations. If there is to be any degree of rehabilitation for these young students, it will necessarily follow that they compensate in some measure for their scholastic deficiency incurred as a result of their mental illness.

The likelihood of their acceptance into and their employment by society progressively declines with retardation in academic progress. Who, in his right mind, would employ a patient for any sort of desirable position if the patient could not read or write or tell time? If the patient is destined for nothing more than custodial care for life, these things become of trifling importance for any environment which is equipped to care for him is equipped to make allowances for his inabilities. If he is to become a responsible citizen, he must meet the basic standards.

Intellectual performance by an individual suffering from a mental condition is almost always impaired, and the encroachment is usually spotty and uneven. The students themselves are unaware of their deficiencies; so it must be up to the teacher to ascertain their lacks (and some of these show up only after considerable time)

and to fill the gaps so that additional increment will be based on a solid structure of accomplishment.

- 2) To provide basic personal skills for the student to use to facilitate his future adjustment.

It seems that one of the first things to go when an individual is besieged by mental illness is his will to make himself a part of cleanliness and neatness. He lets his clothes become dirty, disorderly and ragged; his hair and teeth go unmolested by comb and brush; his school work and letter-writing go to seed; and water seems of little concern to him since a great portion of that which gets on his outside is only by accident.

Since we became a germ-conscious culture some years ago, there has been a compulsion toward neatness which distinguishes our middle class from almost every other subculture on earth. From the point of view of the world outside the hospital a degree of personal neatness not commonly found among hospital populations will be a prerequisite to satisfactory adjustment.

It is recognized that a superficial veneer of tidiness will do little toward amelioration of the basic disorder, but it is felt that gentle urging in that direction, setting examples for the students and the constant recognition of neatness as a symptom of normalcy will do a great deal toward reinstating the student's self-respect when other more pressing problems are resolved.

There is to be considered also the hygienic value of cleanliness and the concomitant morbid mental state that frequently goes with, or is aggravated by, a microgenic disorder; but these are so commonly accepted that they need only be mentioned.

- 3) To provide accredited academic work on primary and secondary levels.

Our class was fortunate in having the coöperation of the Austin Public School System in the accrediting of work done in school; for credit, *per se*, is important to the student as well as to others. For the student it provides some tangible goal for him to strive toward; some milestone which he can use to signify a definite level of achievement to satisfy himself that he is not merely vegetating during his stay in the hospital.

Most establishments which would be prospective employers of the students have no program for testing the proficiency of the

students along these lines; so the symbol of proficiency, the diploma, frequently becomes the universal, critical key to the future.

- 4) To provide information which will serve as a basis for their discrimination in introspection.

It has been noticed that some of the students realize that they have had hallucinations, and the only basis for discrimination between facts and artifacts is their fund of information. The realization that tinnitus is a universal occurrence and not an indication of malady came as a startling revelation to one student. Such understanding of natural phenomenon and the significance of introspective observations will go a long way toward minimizing misinterpretation and should increase the student's grip on reality as he begins to gain some insight into himself.

- 5) To occupy the student constructively.

It has been said, "No wonder epileptics deteriorate: they have nothing to do but sit around all day and watch each other throw fits." In the same vein, residents of mental hospitals who have nothing to do but contemplate their symptoms and vegetate in the behavioral fertilizer of other psychotics can hardly be expected to derive much benefit from their contacts. Educational therapy is designed, at best, to provide goals and incentives to reach those goals—at least, to give the students something to keep them from perseverating on their own morbidity.

- 6) To provide activities and entertainment which are designed to keep the student in contact with his concept of normalcy.

It has been noted that some of the students who gain a bit of insight into their conditions identify themselves with the abnormal. For these students to reach a high degree of normal functioning while isolated in the hospital requires a bootstrap sort of effort. They must identify with normalcy without being exposed to it. The process of deterioration may be nothing more than an internalization of the lax and permissive attitudes of the institution. Abandoning hope for release and realizing that he need have no rigid standards to continue life in the hospital, the student gives up the struggle.

Educational therapy is designed to bridge the gap between normalcy with its intolerance and hospital permissiveness. It

brings the 'outside', non-threateningly, to the student for his acceptance.

7) To provide opportunities for social interaction.

In addition to identifying personally with normalcy an important goal is that of identifying with society as a collection of individuals.

The pseudotolerance which is often exhibited by mental patients toward the behavioral peculiarities of others is not based on acceptance of them, but rather on a wide separation of their respective inner worlds. When these separate worlds of action are brought into juxtaposition, it is usually observed that they react in a characteristic short-sighted, egocentric manner.

If the students could be brought to some degree of realization that one can make allowances for others without loss of status all would be well; but if this is impossible, it might still serve a valuable purpose if one were taught merely to inhibit one's aggressions, thus obviating the vicious cycle of threat and ego defense. It was found possible to use the rapport which was established to circumvent some threatening unpleasantness. It is felt that as *esprit de corps* develops that social pressures will come to mean more in shaping their activities.

In order to develop this feeling, activities which set the students apart from the rest of the hospital population were planned. They included movies on a wide variety of subjects, student-selected field trips such as the museum and Barton Springs (one student caused a furor when he announced to the life guard that he was nine feet tall, weighed three hundred and ten pounds and was the best looking man on the grounds).

There is also in class an opportunity for students of opposite sexes to meet in supervised but permissive atmosphere. Such constitutes experience which should be in the repertoire of the student before he reenters society.

8) To provide opportunities for success.

It is so important that the students be made to feel successful in their activities that rather than being merely a technique for achieving other aims, this is in itself a goal. There is an effort made not to assign any task which is beyond the realm of attainment for the student. He is given no deadlines for its completion and much commendation is rendered when it is finished. If it is

partly correct, it is considered correct with an explanation of the alterations necessary. If it is incorrect in its entirety, no blame or censure or disapproval is indicated, especially when the student seems to be chastising himself, and encouragement and reassurance is always abundant.

"Grades" are optional with the student. Some prefer them and others do not, but in all cases a pittance of work nets a superlative grade. Seldom are they less than 95, and when work is corrected only the second mark, 100 is given.

It has been noticed that these students frequently prefer dull, routine drill. This seems to be a striving for adequacy and security, and it must be catered to before the desire on the part of others for a greater academic achievement. It is hoped that when they feel a genuine security the students will show spontaneous tendencies toward exploration of unknown fields and diversification.

9) To provide an accepting atmosphere.

An accepting atmosphere is indispensable to rapport, to a feeling of non-threat, and hence to a facile learning situation. The key to this seems to be in quiet, interested, unemotional receiving of any novel and startling innovation in student behavior. All the eye-brow-raising details are greeted without panic, and they are referred to thereafter only by the students. Quite naturally, in this population there is a limit beyond which aggression and negativism cannot be tolerated, but, happily, these limits are seldom reached.

On 'those days' students are allowed to read comics, just sit and stare, return to the wards, sleep under the tables or whatever they feel like doing. They are asked frequently if they have suggestions and practically any spontaneous activity is included in the curriculum.

Competition, whether it be between the students or between school and students, is not sponsored. Each student works on his particular assignment and is paced by his own ability rather than any frantic sense of academic inferiority-avoidance which is so frequently used to motivate normal students.

To increase the atmosphere of the school and diminish the hospital influence the students are referred to as such and never as patients.

At first the students seemed to get nowhere and some used this

permissiveness as a loading device throughout, but most seemed to do as well as their limitations would allow.

- 10) To collect data which may shed light on these conditions, their amelioration, education and perhaps their prevention.

In all the literature on the education of exceptional individuals there seems to be a paucity concerning mental disorders, perhaps due to lack of contact with truly abnormal people in the average educational environment. Hypotheses which have been made and on which therapy has been based can be evaluated only by comparisons yielded from this and comparable classes; so for this our class supplied a daily log of events and attitudes, a collection of their work and other miscellaneous items: deposits for the future.

RESULTS

The class met on one hundred days from April through August, and had enrolled eleven students whose attendance ranged from four to ninety-six days. One boy who attended class only one day is not counted.

Academic progress was generally better than had been anticipated. The most outstanding advance was made in typing. The student progressed from an incomplete knowledge of the keyboard to long paragraphs of mistake-free transcription done rhythmically and with promise of speed.

Others showed considerably less improvement. One had hope of mastering the Arabic numeral system from 1 to 12 preparatory to learning to tell time. He managed nine numbers in only one hundred days which was in reality an accomplishment.

Five students made notable progress, either in new subjects or in recapturing departed abilities. One who showed great promise left the warm and mothering confines after only four days of attendance.

Five others showed somewhat less progress, two of these were the lobotomies, one was on shock treatment, and one was the brain damage case. The two aments failed to grow in academic grace, but they did not shrink either.

The personality progress of these students is harder to ascertain. An effective appraisal would necessitate frequent observations by one trained for that purpose, but lacking these facilities, the only alternative was to base the progress reports on the daily contacts

in the school. Five made strides toward gaining self-confidence, socialization and development of self-control. The two who were aments were never a personality problem. The organic case was mostly uninfluenced—as might be supposed. Three showed no improvement, but, happily, no deterioration either.

FUTURE CLASSES

One group omitted from the experimental program would seem to merit special attention. Here in the South Negroes generally maintain an average intellectual achievement lower than that of whites. It would often be an unrewarding task to try to bring the educational average of the Negroes up to the level of the whites, for if a Negro were returned to his peer group with an educational level which was very far in excess of the group he probably would find acceptance by even his own group more difficult—and acceptance by the whites still impossible—thus aggravating the very problem which is crying for solution.

Luckily, the absolute level of education is of much less importance than the relative level. If he returned with a new ability—perhaps the ability to read where before he was illiterate—he would be in a position to command increased respect from his associates and still be accepted.

It would seem clear that the ultimate goals of each case would be based on a thorough study of the case itself and how it articulates with its environment, rather than a remote theoretical criterion.

LIFE ACTIVITIES

The feeling of accomplishment for anyone might also go a long way toward counteracting the stigma of being a mental patient which is all too often still a part of present-day thinking. The patient who is forced to admit, even though only to himself, that he is handicapped in his living because of losses due to his mental illness, is placed at a distinct disadvantage; but if he could say, "Look! I can perform better because of my stay in the hospital," he would be better able to face the readjustment to society and the acceptance of his own illness.

Many things which the non-institutionalized person takes for granted could be the source of acute embarrassment to one who has been out of contact with society for some time. For example,

what embarrassment succeeds a frustrating encounter with the ubiquitous coin machine which even the little children can operate? Likewise the cataclysms which could be caused by the eye operated doors, help-yourself supermarkets, late model automobiles, multi-directional traffic signals, express busses, self-operated elevators, perhaps even dial telephones and the other myriads of gadgets which characterize our age.

For one engaged in the herculean task of readjusting to a life that has passed him by, many such problems might be critical in causing existence either to be smooth and flowing or a succession of accute discomforts.

Toward the end of his stay classes would include 'Living' which stresses smooth manipulation of common situations. The intensity of this phase would depend on the total time he has been effectively removed from society, the state he had reached in his pre-morbid social development and social expectancy.

ORGANIZATION

The inclusion of Educational Therapy would not eliminate such therapies as Occupational, Recreational and Music, nor would it merely superimpose another one of these on the existing structure. Rather it would integrate them into a single, coördinated, multi-directional therapy for a concentrated attack on the illness. Educational Therapy should include all these and anything else that would tend to increase the individual's ability to perform successfully in life.

Because of the variety of backgrounds and the differences in goals of those patients admitted to the Educational program it cannot take on a linear form such as is found in ordinary schools. These assume that everyone begins at the same level and ends at the same level. Here, however, an areal pattern would obtain which would allow one to start at his own level and work upward through his stay in the hospital toward individual goals.

Upon entry into the hospital the usual legal, medical and psychological processing would take place as rapidly as possible. Part of this processing would include information concerning his academic past to assist in his placement. Also included would be items which would shed light on his pre-morbid educational expectations to be used as aids in fixing his goals, and they would be influenced

also by his mental capacity, his degree of adjustment, rate of improvement and social expectancy.

His first classes would be simple, small, informal, permissive and homogenous groups (perhaps cleanliness only to accompany shock treatment). His success would be the desideratum. Classes should include as much variety as possible and still be within easy reach of the student. Their periods should be short, not more than thirty minutes, to accommodate their restricted attention spans. Periodically some physical activity must be provided to assuage the lethargy in the withdrawn ones and to give release to the hyperactive ones; also it has been observed that movies of almost any kind have a general appeal, almost magical, and good results have been experienced with having refreshments (cookies, pop, graham crackers, fruit juices, etc.) at odd times during the day. One half day of classes should be considered a full load at the beginning and as a student's disturbance abates there would be a corresponding increase in the academic load. Students in the experimental class responded well to the idea of having free access to the equipment and being placed on their honor as to its use.

Although permissiveness is the key to the first courses, it must go hand-in-hand with persuasion and a great deal of encouragement. It must be remembered that, for the patient, this is not his major battle; so he may be expected to lack a great deal of enthusiasm for school. These first classes must be limited to six or eight, and even fewer if they are severely disturbed or highly manic.

As the student progresses and is better able to make successful adjustment to additional responsibilities, he will be promoted to a larger class with more exacting standards. This graded permissiveness insures the warmth necessary to the beginner, and gradually 'toughens' him so that he will not be overwhelmed upon release from the hospital.

Concurrently, of course, the psycho-environmental and medical aspects would be operative while treatment of his milieu would be progressing behind the scene. Medicine has long believed that medication followed by reinfection was effort wasted. Similarly here. The social services would have as one of its chief functions either reëducation of the patient's former environment to a point where it no longer constituted a threat to him, or placing the student where he can be accepted. Graduation would be synonymous

with discharge from the hospital, and a diploma should indicate work accomplished.

TEACHER

The teachers for this program will constitute the major problem. If the endeavor is ever to be successful, it must be undertaken completely—which includes an adequate student/therapist ratio. The teachers for such a program would be well-grounded in psychology and the basic concepts of psychotherapy. When all is said and done it is still an organization devoted to the restoration of mental health; education here is merely the key to this restoration, and nothing should obscure this salient fact. The teacher should be acquainted with educational procedures, especially those which deal with the education of exceptional persons.

These two factors alone are not enough. Especially for the first classes, the therapist must have a warm, easy-going, 'mushy' personality which will allow him to identify easily with his students and they with him. Students respond much better to a bit of horizontal interest than to a business like, rigorous, discipline-based authority. They seem to accomplish more in less time when allowed to spend some time in ventilating their feelings in class.

SUMMARY

This is a report on an experimental class in Education Therapy which was conducted at the Austin State Hospital in 1951. The goals of the program were (1) to shorten the stay in the hospital of all who are remissable, (2) to provide a stable viaduct back to normalcy facilitating readjustment, (3) thereby reducing the number of readmissions and (4) eventually reducing the over-all number of mental patients in hospitals.

The chief tool in the task of preparation and readjustment would be an integrated educational program. The aims and purposes of this academic situation in particular, and perhaps for Educational Therapy in general are:

- 1) To provide basic academic skills for the student to use to facilitate his future adjustment.
- 2) To provide basic personal skills for the student to use to facilitate his future adjustment.
- 3) To provide accredited academic work for primary and secondary levels.

- 4) To provide information which will serve as a basis for their discrimination in introspection.
- 5) To occupy the students constructively.
- 6) To provide activities and entertainment which are designed to keep the student in contact with his concept of normalcy.
- 7) To provide opportunities for social interaction.
- 8) To provide opportunities for success.
- 9) To provide an accepting atmosphere.
- 10) To collect data which may shed light on these conditions, their amelioration, education and perhaps their prevention.

WHO IS PENALIZED BY THE PENALTY FOR GUESSING?

ALEX C. SHERRIFFS and DONALD S. BOOMER

University of California

Almost since the introduction of objective tests into the field of educational measurement, teachers have been concerned with the guessing factor as it enters into test scores. Correction formulae of varying degrees of complexity have been developed. All are variations of the original method of subtracting the number wrong from the number right, ignoring omitted items, and considering the remainder as the significant measure.

Generally it has been shown that the application of correction formulae results in a wider range of test scores and in downward adjustment of average scores. Some have claimed that chance effects are reduced and that a more reliable set of scores results (5). Others, however, believe that a correction formula lowers reliability (4, 6). Ruch (7), after surveying the literature to 1929, concludes that the effect of a correction term on reliability is debatable.

All in all, though the effects of the theoretical corrections for chance are completely unclear, such corrections are applied by many teachers in the classroom situation. One extreme defender of the penalty for guessing aims his correction at the 'gamblers', 'bluffers', and 'bullies' among the students, stating that "The results (of penalizing wrong answers) are always the same. Students gradually acquire a new regard for truth, precision of learning, the impropriety of guessing, the shame of bluffing, *et cetera*" (3).

The present writers have for some time believed that the usual notions about the effects of right-minus-wrong scoring have ignored some important factors. Our own reasoning has been somewhat as follows.

A course examination is administered in order to evaluate the present 'achievement' of students—their knowledge of course content, or their ability to utilize their course experience in relevant problem situations.

It is our contention that course examinations should not serve as direct tests of intelligence. (If measurement of intelligence is

desired, quantities of standardized tests for the purpose are available.) Neither ought they be measures of the student's personality, nor should successful performance in the testing situation be contaminated any more than necessary by personality variables.¹

We believe that when the right-minus-wrong instructions are added to an ordinary examination a new set of decisions faces the individual student. For example, such questions must necessarily arise as: "How certain am I of this answer? What is the probability of my being right? If the probability is 80 to 20 that I am right, then am I better off to 'guess' or to omit the item?" Ability to 'figure the odds', subjective levels of confidence in own judgment, and willingness to gamble or to take a chance, are among the variables which might well be expected to influence the right-minus-wrong measure of achievement. Such influence can only make the test less valid as a measure of achievement.

It seems obvious that such a contaminated measure is not only less useful for its avowed purpose, but also that it will work hardships on individual students because of individual differences in shrewdness in test-taking or in other personal factors.

The present study represents an attempt to find evidence as to the presence of contaminating factors in the conventional right-minus-wrong test score.

PROCEDURE

Two approaches to the question under consideration suggest themselves. One approach involves scoring the performance of each student on one test in two ways—by applying a penalty for guessing, and by a simple count of the number of items answered correctly. The two resulting sets of scores can then be compared in their relation to each other and to other pertinent variables. A second approach would involve having some course examinations of the right-minus-wrong type and others of the ordinary 'no

¹ We agree with Cronbach (1) that "When the test is used to fill other purposes, the score becomes invalid for the primary purpose of measurement; essentially the more elements that enter a score, the less meaningful it becomes." A few writers have considered the dangers of contaminating measures of course achievement with the effects of personality differences (8, 10, 13) but only Votaw (11) has approached this question experimentally. He presents data on the relationship of measures of dominance-submission to guessing behavior.

penalty' variety; then comparing the student's scores on the two types of tests.

There are advantages and disadvantages inherent in both designs. The first approach was followed in this investigation.

Subjects.—The subjects were students in the second semester course in psychology designed for non-psychology majors at the University of California. Of the four hundred and fifty students who were enrolled in this course three hundred and ninety are active subjects in this investigation. The criteria for inclusion were that the students have participated in four midterms and the final examination and that the students have satisfactorily completed the Minnesota Multiphasic Psychological Inventory.

Experimental Examination.—The first midterm examination consisting of fifty true-false items served as the experimental examination for this study. The class was divided between two examination rooms, students whose last names fell in the first five letters of the alphabet going to one room, the remainder to the other.

In one room ninety-eight subjects took the examination on a 'no penalty for guessing' basis. "Answer all questions, please." This group served as a control for the test-taking condition.

In the other room two hundred and ninety-two subjects were told that the examination was to be scored on a right-minus-wrong basis with no score for omitted items. After these students had completed their tests they were instructed to circle all the items they had omitted. Following the completion of this task, they were asked now to indicate what they believed to be the most probable right answers for the previously omitted items. The circling technique made it possible to obtain three scores for each student's performance on the examination. These were (1) a right-minus-wrong score ($R-W$), (2) a score representing the total number of items that the student answered correctly when required to try all (No. Right), and (3) a penalty score (Penalty Score) determined for each student by the difference between his $R-W$ score and his No. Right score.

Criteria for course achievement.—To aid in interpretation of results some definition of course achievement was desired. A best approximation was utilized for the study, consisting of the combined scores on the three midterms following the experimental examination and the final examination. True-false questions,

multiple-choice questions and essay questions each contributed about one-third to this measure. This measure of course achievement turned out to be stable, to yield a wide distribution of scores for our sample of subjects, and, in terms of subjective evaluations of staff, generally to rank students in accord with expectations based upon office hour contacts.

Personality dimension.—It was our hypothesis that students who lacked self-confidence, especially confidence in their own judgment, who had difficulties in making decisions, and who were easily threatened by ambiguous situations, would be penalized by the R — W test condition. To test this hypothesis we desired a personality measure which might be expected to relate to some of these variables.

The measure determined upon was the A scale of the MMPI. This is a new scale of thirty-nine items developed by Welsh (12). It is claimed that people scoring high on this scale are characterized by introversion, rumination, anxiety, low self-esteem, and undue concern with the impression they make on others. The 'flavor' of the scale, its face-validity, is conveyed by such items as:

Criticism or scolding hurts me terribly (true)

I worry quite a bit over possible misfortunes (true)

I have often lost out on things because I couldn't make up my mind soon enough (true)

The group form of the MMPI was administered to the entire class in mid-semester, and for the purpose of this investigation was scored for the 'A scale'.²

RESULTS

Experimental Groups

1) 'R — W' and 'No Right' as measures of achievement.—Comparisons were made of the two scoring methods applied to the fifty-point experimental midterm. The Pearson Product Moment correlation between the 'R — W' and 'No Rt' measures was .96.

² The investigators did not believe, of course, that the 'A scale' was the only measure worth following up, nor did they believe that they knew at the present stage of its validation exactly what aspects of personality it measured. However, available evidence about this scale suggested it to be a good bet. Insofar as the case against 'R-W' scoring goes, contamination by any personality factors, whether or not closely defined, would be important.

The correlation between the 'R - W' scoring of the experimental midterm and the criterion measure of course achievement was .48; between 'No Rt' and the criterion the correlation was .49.

The 'R - W' scoring method yielded a lower mean score and a more variable distribution of scores. However, this must be seen in relation to the extremely high correlation between 'R - W' and 'No Rt' scoring, and essentially the same degree of relationship between these scoring methods and the criterion for course achievement.

2) *The method of scoring and the personality measure.*—The 'A scale' yielded a skewed distribution of scores with a mean of 9.93, a standard deviation of 7.8, and a range from 0 to 35.

Correlation coefficients were calculated in an effort to assess the degree of relationship between the 'A scale' score and the two scoring approaches to the experimental midterm. The correlation between 'A' and 'R - W' was $-.23$; between 'A' and 'No Rt' it was $-.22$. A coefficient was also derived for the relationship between the 'A scale' and the achievement criterion. This correlation was $-.20$. These correlations are all significantly different from zero.³

The above findings do not answer the question as to whether individual students were differentially handicapped by 'R - W' scoring because of personality factors. A comparison of those students high in maladjustment as indicated by the 'A scale' with those low should, however, offer crucial evidence. It was decided to study approximately fifty cases at each end of the distribution. Because of the presence of a relationship between the personality variable and course and examination performance, subjects from the two ends of the distribution were paired on the basis of their total course achievement. It was possible to obtain forty-seven pairs, matched within two points on the achievement criterion. Thus we are comparing two groups of students who functioned equally well on the accepted measure of course mastery and who differ only in that they are the 'highs' and the 'lows' on the personality variable in question. How these groups functioned on the experimental midterm is summarized in Table I.

³ Examination of the scatter plots reveals the relationship between the 'A scale' and all other measures to be curvilinear. Thus the reported correlations can only be used to predict in one direction—from the 'A scale' to the performance measures.

TABLE I.—COMPARISON OF PERFORMANCE ON EXPERIMENTAL MIDTERM OF STUDENTS WITH 'A SCALE' SCORES OF 15 AND OVER WITH THOSE WITH SCORES OF 4 AND LESS

	<i>Highs</i> (A scale score of 15 and greater)	<i>Lows</i> (A scale score of 4 and less)	Mean Diff.	^a Std. Error	CR
Mean R-W Score	18.87	22.85	3.98	1.23	3.23
Mean No. Rt. Score	34.89	36.32	1.43	.89	1.61
Mean <i>Penalty</i> Score*	16.02	13.47	2.55	.837	3.05

* The *penalty* score represents the difference between the gross scores on R-W and on No. Rt. It is not a completely satisfactory statistic. However, the skewness of the R-W distribution made questionable the use of differences between standard scores.

These findings show that students with high 'A' scores were penalized by the use of 'R - W' scoring in comparison with students whose 'A scale' scores were low. The difference between these groups is statistically significant at better than the .01 level. Further, this is a difference of practical significance, amounting to almost a full letter grade within the 'R - W' distribution.

3) *Analysis of patterns of performance under the R - W condition.*—A study of the 'omitting' behavior of the subjects of the experimental group was made in an effort to throw additional light on the manner in which the higher 'A scale' student was penalized.

One hundred and eighty-seven students omitted three or more items. These subjects were divided into two groups: those whose 'A scale' scores were 15 or over and those with scores of 14 and less. Next, these subjects were paired, high 'A' with low 'A' on the basis of their 'No Rt' scores. This was to match them in terms of their knowledge of the material covered by the examination. It was possible to obtain forty-four perfectly matched pairs.

Two comparisons were then made. The first was in terms of the number of items omitted by the two groups. Students with higher 'A scale' scores omitted under the 'R - W' conditions 6.25 items, on the average; those with low 'A scale' scores omitted 5.09. This difference results in a critical ratio of 2.03 and is significant at the .05 level.

A second comparison was made. The proportion of items later ~~answered~~ correctly (after circling) by the two groups was studied

to see whether the higher 'A' subjects differed from the low 'A' subjects in this regard. Sixty-three per cent of the items omitted by the higher 'A' subjects were later answered correctly by them. Fifty per cent of omitted items were later answered correctly by the low 'A' subjects. The significance of this difference of thirteen per cent was assessed in three ways: In terms of the t resulting from analysis of correlated proportions, in terms of the difference between paired percentages utilizing the $\sqrt{\text{arc-sine}}$ transformation, and according to the non-parametric 'Sign Test' suggested by Dixon and Mood (2). In all three cases this difference proved significant at the .01 level or better.

These findings reveal that the excess penalty which we have found to accrue to more maladjusted students comes in part, at least, from two factors. The more maladjusted students omit more items, and they omit a larger proportion of items that they are capable of answering.⁴

Control Group

It will be recalled that ninety-six of our subjects took the identical midterm as did the experimental group, but under different conditions. The controls never believed the examination to be of the penalty-for-guessing type, taking it only on the 'try all the questions' basis.

The control group data provides information making possible sharper interpretation of certain of the experimental group results.

1) Did the experimental group subjects try less hard on the previously omitted (circled) items when instructed to answer them than they would have done had the examination been a conventional 'number right' type in the first place?

2) Did the very situation of being confronted with a 'R - W' test modify the performances of students in general?

⁴ Additional information was obtained on this matter by taking another slice at the data. Sixty-six pairs of students were studied. These were matched for 'No. Rt.', but divided into two groups: those who answered correctly better than fifty per cent of those they omitted, and those who answered correctly only fifty per cent or less. A comparison of the 'A scale' scores for those two groups revealed a mean 'A scale' score for those students more successful on omitted items of 11.73, and a mean score for those less successful of 8.42. The mean difference of 3.31 points has a standard error of 1.25 points. The Critical Ratio for this difference is 2.65. The difference, then, is significant at the .01 level.

TABLE 2.—COMPARISON OF CONTROL AND EXPERIMENTAL GROUPS ON BASIC VARIABLES

	Achievement Criterion		No. Rt. on Experimental Exam ³		A scale Scores	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Control subjects (n = 96)	242.97	20.61	35.50	4.40	10.07	8.51
Experimental subjects (n = 292)	241.87	19.59	35.87	4.21	9.95	7.80

3) In the case of those students who had been found to be handicapped by the penalty-for-guessing scoring, was part of the explanation of this handicap to be found in the general atmosphere of the test situation itself?

Should the control group turn in a better performance on the examination than did the experimental group, the answer to any or all of the above questions could be 'yes'. It would be difficult with our data to determine the relative contributions of carelessness and of insecurity. However, should there prove to be no difference, then the indications would be negative on both counts.

A comparison of the control and experimental groups follows.

None of these differences between the Experimental and Control groups was significant at better than the .30 level.

In the case of the control group a correlation of .56 was found between the experimental midterm and the criterion for achievement. When this is compared with the .49 in the case of the experimental group, the difference is not significant.⁵

Two final checks were made. Those subjects of the control group who had 'A scale' scores of 15 and over were compared in their performance with those who had scores of 14 and less. Ad-

⁵ The correlation between the 'A scale' and the control group scores on the experimental midterm was $-.003$. Between the 'A scale' and the criterion for achievement it was $-.070$.

The significance of the difference between these 'zero' control correlations and the comparable experimental ones (of the order of $-.20$) was calculated utilizing Fisher's Transformation. These differences do reach the .05 level of significance. We are unable to find an explanation for these differences.

ditionally, control subjects were matched for 'A scale' score with cases from the experimental group. All other variables were disregarded during the pairing. No significant difference resulted from any comparison. High 'A' subjects in the control group were not different from low 'A' subjects in regard to 'No Rt' on the experimental midterm, or in regard to course achievement. High 'A' subjects in the control group were not different from high 'A' subjects in the experimental group, nor were low 'A' controls different from low 'A' experimental subjects.

Study of the control group, then, leads us to the conclusion that there is no evidence of carelessness, or of disturbance in 'No Rt' performance simply from being in the 'R - W' testing situation.

CONCLUSIONS

On the basis of this investigation we find no reason to reject our original hypothesis. We find that students who score on the maladjusted end of the 'A scale' are handicapped by being scored on a 'R - W' basis. This holds true even when the students' knowledge of course material is held constant. The evidence suggests that these students are penalized by their tendency to omit more items, and to omit items the answers to which they know.

We can see no justification for teachers continuing to use corrections for guessing. The assets accruing to such a system have always been equivocal, and the liabilities, we believe, are now apparent.

We hope that Traxler (9) is unduly pessimistic when he says that he doubts that school people will ever be willing to encourage pupils to guess at every item in a test. He believes that these educators would feel this to be condoning carelessness and loose thinking. We hope that the educator will also take into account principles of fairness, and of respect for the 'whole child'.

REFERENCES

- 1) L. J. Cronbach, "True-false tests: A reply to Count Etoxinod." *Education*, 1941, 62, 59-61.
- 2) W. J. Dixon and A. M. Mood, "The statistical sign test." *J. Amer. Stat. Assn.*, 1946, 41, 557-566.
- 3) S. Etoxinod, "How to checkmate certain vicious consequences of true-false tests." *Education*, 1940, 61, 223-227.
- 4) Harry A. Greene and A. N. Jorgensen, *The Use and Interpretation of High-School Tests*. New York: Longmans, Green & Co., 1935.

- 5) Herbert E. Hawkes, E. F. Linn and C. R. Mann. *The Construction and Use of Achievement Examinations*. Boston: Houghton-Mifflin Co., 1936.
- 6) C. C. Moore, "Rights minus wrongs method of correcting chance factors in the true-false examination." *Pedagogical Seminary*, 1940, 57, 317-326.
- 7) G. M. Ruch, *The Objective or New Type-Examination*. Chicago: Scott, Foresman & Co., 1929.
- 8) Frances Swineford, "The measurement of a personality trait." *J. Educ. Psychol.*, 1938, 29, 295-300.
- 9) J. W. Varty, "Guessing on examinations—Is it worth while?" *Educ. Forum*, 1946, 10, 205-212.
- 10) D. F. Votaw. "The effect of do-not-guess directions on the validity of true-false or multiple-choice tests." *J. Educ. Psychol.*, 1936, 698-703.
- 11) George S. Welsh. Personal communication to writers.
- 12) Llewellyn N. Wiley and Otis C. Trimble, "The ordinary objective test as a possible criterion of certain personality traits." *School and Society*, 1936, 43, 446-448.

TOPOGRAPHICAL ORIENTATION IN NAVAL AVIATION CADETS¹

BRANT CLARK AND R. DANIEL MALONE

United States Naval School of Aviation Medicine, Pensacola, Florida

The problem of how an individual orients himself to geographical points is an old one and has involved much theoretical controversy. Nevertheless, very little clear understanding of the specific processes involved has resulted, although radical postulations of 'special senses' in both animals and man have been common. The most intensive studies have been made by educators who have been interested in geography (4, 15, 16, 18, 19, 23, 27). The majority of these studies have investigated childrens' ability to point north or to other compass points, while some of them have been concerned with the childrens' ability to indicate the direction of places within their own community as well as distant points. Only a half dozen studies have been found which were concerned with topographical orientation in human adults. Two of these papers were essentially anecdotal in nature and merely describe the activities of a limited number of individuals having either a high degree of spatial orientation or experiences of being seriously disoriented (2, 3). Only two investigations (14, 31) have employed a large number of subjects. These studies have demonstrated wide individual differences in orientation to geographical points as well as constant group errors. It has been demonstrated that scores on tests of orientation to geographical points show low correlations with general intelligence, age, knowledge of the vicinity, sex and school experience. Vision and mental maps appear to play a pre-dominate rôle in aiding the individual in orientation.

The importance of adequate orientation is well-known to pilots of aircraft (7, 8, 22, 26, 28). This has been demonstrated by analyses of flight safety records, (7, 26, 28) analyses of the causes of failures in flight training, (17) and the relationship between spatial and visual orientation to success in flight training (17, 20). Experiences reported by pilots (28) show that in spite of their very

¹ Opinions or conclusions in this report are those of the authors. They are not to be construed as necessarily reflecting the view or the endorsement of the Navy Department.

efficient navigational aids, pilots do occasionally become confused with regard to their geographical position. It is significant to note that when this disorientation occurs, it is associated with marked nervousness and uncertainty on the part of the individual (2, 28).

It was the purpose of this study: 1) to determine how well Naval Aviation cadets are oriented to geographical points; 2) to identify common errors in orientation; and 3) to compare topographical orientation with two other tests of spatial orientation, tests of other mental abilities, and certain selected personal data.

PROCEDURE

The primary, methodological problem was to develop a technique which would determine how accurately Naval Aviation cadets are oriented to certain geographical points, since no standardized tests were available. A modification of the pointing procedure previously devised by High (14) was used because it was simple, involved a minimum of extraneous factors, was readily adaptable to group testing, and the results could be treated quantitatively.

The test, which was finally developed, consisted of two parts containing twenty items each. Each part in turn consisted of a list of ten well-known cities in the United States and ten well-known foreign cities (Tables I and II). In Part I, the subjects were asked to imagine that they were in Chicago, Illinois. Using a compass rose as a reference, they were instructed to indicate the direction they would travel from Chicago to each of the twenty cities by the shortest possible route. The directions were: "Write in Column C below the direction you would go from *Chicago, Illinois*, to reach each city by the shortest possible route. Refer to the compass rose for the directions." They read the heading from the compass rose in degrees and recorded it in the appropriate column on the test sheet. The directions for Part II were identical with those for Part I, except that the point of origin was Pensacola, Florida. These cities were selected on the basis of High's study and a preliminary study at Pensacola, eight of them being common to both points of origin. For each part, the cities were selected so that they were well distributed around the compass rose. All of the subjects were asked to face north during the test because the direction one faces during such a test has been found to be of importance in earlier studies (14, 16).

The test was designed to obtain data for the analysis of errors on individual cities and to give a total orientation score. The scoring procedure used to obtain the total score on this test was somewhat unconventional. First, the latitude and longitude of each of the cities were obtained from Goode's atlas (6) for computational purposes. Second, initial great circle headings for each city were calculated using a method devised by Ageton (1). This heading is the direction a person would take to go from one point on the earth's surface to another by the 'shortest possible route'. Such headings are used by ships and aircraft when great circle sailing is followed to actually use the shortest possible distance between two points on a sphere. Third, the mean and standard deviation of the cadets' estimates for each of the forty cities were calculated. Fourth, keys were made out for each item on the test in terms of one-half sigma deviations from the initial great circle heading which was the correct response. Deviations from this heading within plus or minus one-half sigma were scored zero, and each succeeding one-half sigma step was scored as a unit, the direction of the deviation being disregarded. The sum of these deviations for all of the forty items was then used as the individual's score on the test. This score was an "error score" inasmuch as it was the sum of the individual's deviations from the initial great circle heading for each city.

Mercator maps, which plot the earth as a flat rather than a spherical surface, are commonly used in schools and elsewhere for convenience. Such projections of the earth involve gross errors, especially near the poles. However, they are widely employed so headings using Mercator plots of the cities were also determined for comparative purposes (Tables I and II). For some cities the two headings are only slightly different, but for others the errors are considerable as in the case of Moscow and Rome (Table I).

In order to compare topographical orientation with other psychological traits, additional test scores were obtained from the students' training jackets. Two additional tests of spatial ability were also given to the cadets in addition to the tests of topographical orientation which were administered during the Naval Aviation cadet's indoctrination week. The tests were:

- 1) *Essentials of Mathematics for Flight Students*.—This test was prepared by the Central Examining Board of the Naval Air Training Command. It is a ninety-minute test containing

hundred and five items which varied in content from simple arithmetic computation to the reading of graphs and arithmetic reasoning given to flight students during their indoctrination week. The test is essentially a power test, since the time limit is very liberal. The standard scores obtained routinely on this test and recorded in the training records were used. No data on the reliability of the test are available.

2) *Aviation Classification Test*.—This is a group test of general intelligence used in the original selection of the Naval Aviation cadets, and was designed originally to predict success in ground school training. It consists of one hundred and twelve items measuring vocabulary, comparison, practical judgment, and arithmetic. The testing time is forty-five minutes. The reliability of the test has been found to be approximately .80.

3) *Mechanical Comprehension Test*.—This group test is also used routinely in the original selection of Naval Aviation cadets. It is essentially a test of elementary physics and mechanical relationships consisting of seventy-six items with a time limit of forty-five minutes. It was designed originally as a predictor of success in flight training. The test-retest reliability has been found to be approximately .75.

4) *Test of Correctness and Effectiveness of Expression*.—This is Test I, College Form A, of the United States Armed Forces Institute Tests. There are two parts: The first emphasizes spelling and the second contains material on grammar, style, best words and the like. The time for this test is two hours and the score is the number of items correct. No data are available on the reliability of this test.

5) *American Council on Education, Psychological Examination for College Freshman*, (1947).—Standard scores were obtained from the students' records for Q (Non-language), L (Language) and total score. The standard scores used were based on a group of midshipmen having completed two years of college work.

6) *Spatial Orientation*.—This test which is No. 5 in the Guilford-Zimmerman Aptitude Survey (13), was designed primarily to measure the ability to appreciate spatial relations of things with reference to the human body. The test contains sixty-five multiple-choice items in which the subject is asked to make certain judgments of the position of the prow of a boat with respect to an

horizon and other objects in the visual field. The working time on the test is 10 minutes, and the reliability given in the manual is 0.88.

7. *Spatial Visualization*.—This is No. 6 of the Guilford-Zimmerman Aptitude Survey, and is designed to measure processes of imagining movements, transformations, or other changes in visual objects. The subject is asked to make judgments of the position of a clock which may be rotated about its three major axes. There are sixty-two items with a working time of thirty minutes.

The following personal data were obtained during the regular testing period: 1) age, 2) home city and state, 3) flight experience, 4) educational level, 5) travel in the United States (using a five-point check list), 6) travel outside of the continental limits of the United States (using a five-point check list), 7) geography courses taken within the last three years. In addition to this the cadets were asked to give an estimate of their own orientation using a five-point scale as follows: 8) orientation to the major points of the compass, 9) orientation during travel, 10) attention to the direction of travel in a strange area, 11) use of 'mental maps', 12) visualization of the earth (Table IV).

All of the tests with the exception of the ACT and MCT were given during the first few days at Pre-Flight School. These tests were part of an extensive battery given for experimental use or for use in the Training Command. The data obtained from these tests were coded and punched into IBM cards for computational purposes. The number of individuals used for computational purposes varies somewhat throughout the report since data were not available for all cadets on all tests, however, the specific N's are indicated in each case.

The subjects were two hundred and forty-two Naval Aviation cadets, the majority of whom were entering the Navy from civilian life; the remainder were entering from the enlisted ranks of the Navy and Marine Corps. They had passed the Naval Aviation standards on the Aviation Classification Test (above), a Mechanical Ability Test (above), and a Biographical Inventory. Eighty-five per cent of the group had completed at least two years of college. The age range was from nineteen to twenty-seven years with a mean of 21.57 and a sigma of 1.47 years.

RESULTS

(A) Orientation to Specific Geographical Points

The mean and median estimated headings for the several cities are listed in Tables I and II. The distributions of the cadets' estimates of the twenty cities for each point of origin were typically skewed to a moderate degree. In every case the range of estimated

CITY	INITIAL GREAT CIRCLE HEADING	MERCATOR HEADING	N	MEAN ESTIMATED HEADING	MEDIAN ESTIMATED HEADING	STANDARD DEVIATION	INDEX OF SKEWNESS	MEAN DIFFERENCE (Col. 2-5)	CRITICAL RATIO	MEAN DIFFERENCE (Col. 2-5)	CRITICAL RATIO
New York City	68	98	238	84.74	89.10	14.35	.91	3.26	3.56	13.26	14.26
Pensacola	178	179	233	176.40	168.60	12.40	1.89	1.60	1.98	2.60	3.21
Miami	157	163	240	166.08	155.40	13.38	2.39	9.08	10.53	3.08	3.57
San Diego	258	244	240	253.85	250.32	11.85	.89	4.15	5.42	9.85	12.87
New Orleans	190	189	238	187.20	184.95	12.79	.53	2.80	3.39	1.80	2.16
Seattle	295	289	238	286.52	284.95	11.95	.39	8.48	10.95	2.48	3.23
Reno	256	263	239	263.99	261.70	14.02	.49	7.99	8.81	0.99	1.09
Dallas	222	212	242	215.08	218.79	16.80	.66	6.92	6.38	3.08	2.79
Atlanta	161	166	241	166.30	152.50	20.90	1.98	5.30	3.88	0.30	.16
Portland, Me.	76	80	237	73.17	66.80	14.34	1.33	2.83	3.08	6.83	7.39
London	47	77	238	86.92	81.00	18.60	.95	39.92	33.17	9.92	8.16
Tokyo	323	266	235	265.91	277.60	23.83	1.47	57.09	36.79	0.09	.05
Panama	165	169	240	183.68	186.82	18.65	.49	24.88	20.60	20.12	16.74
Cape Town	106	142	233	132.95	132.70	22.60	.03	24.95	16.79	9.05	6.08
Moscow	29	75	242	16.19	12.30	49.30	.24	12.81	4.05	58.81	18.45
Mexico City	207	201	240	197.29	204.35	16.90	1.26	9.71	8.83	3.71	3.41
Anchorage	320	310	236	325.12	319.02	15.20	1.20	5.12	5.22	15.12	15.15
Rome	52	90	237	100.76	105.05	18.99	.68	46.76	39.55	10.76	8.78
Rio	138	150	238	173.56	170.24	15.94	.62	35.56	34.41	23.56	22.79
Oslo	36	64	232	74.75	64.70	21.10	1.43	38.75	28.02	10.75	7.75

TOPOGRAPHICAL ORIENTATION WITH POINT OF ORIGIN CHICAGO - PART I

Table I

CITY	INITIAL GREAT CIRCLE HEADING	MERCATOR HEADING	N	MEAN ESTIMATED HEADING	MEDIAN ESTIMATED HEADING	STANDARD DEVIATION	INDEX OF SKEWNESS	MEAN DIFFERENCE (Col. 2-5)	CRITICAL RATIO	MEAN DIFFERENCE (Col. 2-5)	CRITICAL RATIO
New York	43	41	239	26.32	23.74	11.97	.86	16.68	21.54	14.68	19.07
Chicago	356	399	238	353.31	351.45	12.84	.43	4.69	5.71	5.69	6.78
Miami	125	132	241	142.50	134.79	18.50	1.26	17.50	14.68	10.50	8.84
New Orleans	260	259	237	274.07	269.25	12.25	1.17	14.07	17.70	15.07	18.96
Atlanta	35	33	235	36.75	39.65	30.15	.28	1.75	.89	3.75	1.84
Denver	308	308	239	301.6	300.23	14.64	.27	6.39	6.75	6.39	6.82
Corpus Christi	256	251	237	260.65	264.15	12.00	.88	4.65	5.96	9.65	12.32
Seattle	313	310	236	305.55	310.94	12.00	1.35	7.45	9.55	4.45	5.70
Cleveland	20	20	240	9.44	6.30	15.72	.60	10.56	10.41	10.56	10.39
Lincoln, Neb.	318	330	238	332.00	323.07	17.55	1.53	14.00	12.31	2.00	1.69
London	43	66	237	72.25	62.55	16.13	1.80	29.25	27.93	6.25	6.01
Tokyo	323	273	237	296.31	288.80	23.47	.96	26.69	17.56	23.31	15.25
Honolulu, T.H.	280	260	234	275.31	274.30	15.78	.19	4.69	4.55	15.31	14.84
Panama	160	162	238	189.92	194.60	24.10	.58	29.92	19.16	27.92	17.86
Mexico City	227	223	241	234.76	234.25	23.25	.06	7.76	4.77	11.76	7.29
Guaymas	131	90	231	90.13	97.80	20.19	1.14	40.87	30.70	.13	.15
Toronto	23	20	241	11.40	13.51	14.50	.45	11.60	12.42	8.60	9.15
Moscow	325	263	235	259.86	267.92	20.43	1.18	65.14	48.97	3.14	2.30
Rapahovik	30	36	237	31.51	32.70	15.06	.24	1.51	1.54	4.49	4.62
Buenos Aires	155	158	237	180.77	180.45	19.10	.05	25.77	20.79	22.77	48.33

TOPOGRAPHICAL ORIENTATION WITH POINT OF ORIGIN PENSACOLA - PART II

Table 2

headings was extensive as has been found with other groups (14, 31). Errors made by individuals approached 180° (i.e., the maximum possible error) in several cases, e.g., Chicago to Cape Town. More accurate means and particularly standard deviations were obtained by eliminating the atypical cases which it was believed were the result of errors in recording or reading. This was done by making an initial distribution for each city, determining the standard deviation and then eliminating all of those judgments which deviated more than three standard deviations above or below the mean. Therefore, the data listed in Tables I and II include only those persons who deviated three standard deviations or less from the mean. The number of eliminations in each case can be determined by subtracting the N in the table from two hundred forty-two, which was the total number of cadets tested. The number of eliminations varied from zero (Chicago to Dallas and Moscow) to eleven (Pensacola to Cairo). The major effect of this procedure was to attenuate the standard deviation while the mean was only slightly affected.

Comparison of cadets' judgments with the calculated initial great circle heading.—Deviations of the group from the initial great circle headings will give a measure of the orientation of the group to these geographical points considering the earth as a sphere. The data are uniform in showing gross individual and group deviations from the true directions (Tables I and II). Significant differences ($C.R. > 3$) between the great circle headings and the group mean headings occurred in thirty-seven out of the forty judgments. The deviations varied from 1.60° (Pensacola) to 57.09° (Tokyo) with the origin at Chicago and from 1.5° (Reykjavik) to 65.1° (Manila) with the origin at Pensacola. The errors were substantially less for points within the continental limits of the United States. For these points, the mean of the judgments deviated from the true direction for different cities by amounts which varied from 1.60° (Chicago-Pensacola) to 17.50° (Pensacola-Miami) with a mean of 7.51° , whereas for cities located abroad, the deviations varied between 1.51° (Pensacola-Reykjavik) to 65.14° (Pensacola-Manila) with a mean of 27.04° . The difference between the mean of the errors within the United States and abroad were significant beyond the .01 level of confidence. The errors to points within the United States were less with the origin at Chicago than when the origin was Pensacola ($P = .05$). On the other hand, there was little dif-

ference in the errors between the two points of origin for cities located abroad. These data show that there are constant tendencies within the cadet group to misjudge the true directions of cities within the United States and abroad. However, these group errors are much less than the errors made by individuals. Some illustrations will be useful in showing the size of these errors: New York is almost due east of Chicago, but several people indicated that it is northeast, while several others indicated that it is southeast of Chicago. From Pensacola to Panama, the initial great circle heading is approximately southeast, while the estimated direction by nearly half of the cadets was southwest. Very large errors were also found from Chicago to Tokyo, Moscow, Rome, and Oslo, and from Pensacola to Cairo, Manila, and Tokyo.

Comparison of cadets' judgments with headings determined from Mercator projection maps.—A major problem which has confronted map-makers for many years is how to make a satisfactory projection of the spherical surface of the earth on a plane surface. All of the methods which have been devised to do this result in certain distortions of shape, scale, angle, and area. The Mercator projection has been widely used in atlases in the past notwithstanding the fact that it produces gross distortions of size as the poles are approached. Deviations of the cadets' judgments from the direction of the cities as determined from Mercator projections were calculated as in the preceding section (Tables I and II). The mean of the deviations of the cadets' judgments from the Mercator heading were substantial. However, these deviations were less than the deviations from the initial great circle headings. Differences between the mean of the cadets' judgments and the Mercator heading between the cities exhibited critical ratios greater than three in thirty-two out of the forty judgments. The mean of the deviations for points within the United States was less for the Mercator projections than for the great circle headings, but this difference was not statistically significant. The deviations from the Mercator headings for cities outside of the continental limits of the United States were significantly less than those for the great circle headings ($P = .02$). The deviations varied from 0.09° (Tokyo) to 23.56° (Rio de Janeiro) with the origin at Chicago and from 0.13° (Cairo) to 27.92° (Panama) with the origin at Pensacola. For cities within the continental limits of the United States, the smallest deviation for the group was 0.30° (Chicago-Atlanta),

while the greatest was 15.07° (Pensacola-New Orleans). The mean deviations from Mercator headings within the United States was significantly less than for cities abroad ($P = .03$). These data show clearly that gross deviations from the Mercator headings occur both for the group and for individuals within the group. It is also evident that the cadets are more closely oriented to the earth as a flat surface than to the earth as a sphere.

Discussion.—These results demonstrate that many Naval Aviation cadets are very poorly oriented to certain geographical points. They also show that there are large group errors as well as individual errors. The results are in general agreement with other studies which have found great variability from person to person, and substantial variation from place to place with regard to group errors (14, 31, 32). However, these results are directly opposed to Woodring's (32) reports of bimodal distributions. His subjects appeared to fall into two groups, one of which was well oriented and the other making approximately a chance score. No clearly defined bimodal distributions were found in the forty cities studied here, although the distributions were skewed. The explanation of this difference is to be found in the fact that the cadet population was larger and much more homogenous than the group studied by Woodring. These cadets were all men of a limited age range, having high aptitude test scores and, in general, extensive academic background. On the other hand, Woodring's group was highly heterogeneous. "Of the one hundred fifty-six subjects, one hundred and seven were white and forty-nine were Negroes; one hundred and seventeen were males and thirty-nine were females. The age range was from seventeen-sixty. Intelligence quotient—by the Revised Stanford Binet Scale—ranged from 44-122. The educational range was from no formal education to college graduation." (32) Since Woodring found several of these factors to be related to orientation, it is conceivable that selective factors may have contributed to the bimodal distribution in his population.

The results of this study are in agreement with High's (14) finding that foreign cities show greater errors than cities within the United States and Woodring's finding that distant cities show greater errors. It is interesting to note, however, that High (14) found great variability for points within the city where his subjects took the test. This may have been a function of the selection of his points and the fact that some of his subjects were not thor-

oughly familiar with the city. However, High's data suggest that a person may be poorly oriented in a limited area with which he is fairly familiar but well oriented in a larger area. Additional comparisons with High's data are possible since ten of his cities were identical with the ones used in this study for the point of origin at Chicago. The cadets were somewhat more accurate and less variable than the college group. However, the differences between the groups do not appear to be great in the light of the differences between the groups and the test situations. There was no significant difference ($C.R. < 3$) for four of the cities (New York, Panama, Cape Town, Moscow) while significant differences were found for six cities (Miami, New Orleans, London, Tokyo, San Diego, and Seattle).

The data show that the orientation of the cadets at the beginning of training is much closer to Mercator projections than to great circle or global projections. This is true in spite of the fact that the deviations are large in both cases. It would appear that the wide use of Mercator and other similar flat projection maps in school and elsewhere has had a marked effect on the cadets' orientation to points on the earth producing marked disorientation, particularly to foreign cities. Greater emphasis on the relation of geographical points to each other on a globe in schools and elsewhere would appear highly profitable to reduce or eliminate such gross errors.

No attempt was made in this study to analyze the mechanism used to make the judgments of the direction of the cities. However, certain inferences may be drawn from the data at hand in terms of the type of errors made. Firstly, it is evident that the cadets as a group have acquired certain erroneous associations in terms of distances and/or direction. For example, the cadets as a group believe that Panama is well west of Chicago and Pensacola when it is actually east of these cities. Secondly, the cadets' orientation is much closer to Mercator projections than to a global map. Indeed, they show little appreciation of great circle courses between points and apparently have little appreciation of the location of cities on the world as a sphere. An illustration of the very large errors which result from this type of disorientation is to be found in the group deviation of 65.14° from the great circle heading from Pensacola to Manila. Although Manila is slightly south of Pensacola, the great circle heading would be roughly north-northwest because the

two cities are well north of the equator. However, the mean of the cadets' judgments for the shortest route was slightly south of due west (Table II). The deviation of this direction from the Mercator heading was only 3.14° which was not significant statistically. In other words, the cadets as a group are fairly well oriented to Manila from the point of view of Mercator projections, but they are thoroughly disoriented from a global point of view. Similar gross deviations from great circle headings with small deviations from Mercator headings are found from Pensacola to Cairo and London and from Chicago to London, Rome and Tokyo (Tables I and II). It is interesting to note that the deviations from the Mercator heading are greater from Chicago to Moscow. The cadets tend to indicate a heading close to the North. This may be explained as an artifact resulting from the instructions which point out that the Great Siberian Plain can be reached most directly by passing near the North Pole. These results give an explanation for the fact that foreign cities show greater errors than cities within the United States. Mercator projections are sufficiently accurate within the United States to produce only small errors, whereas Mercator projections are inadequate for cities abroad. Foreign cities like Toronto and Mexico City which are fairly well represented by the Mercator projection show errors of similar amounts for both Mercator projection and great circle heading.

(B) The Relationship of Topographical Orientation to Other Psychological Factors

The intercorrelations between the scores on the topographical orientation test and other psychological factors are summarized in Table III. Most of these correlations are not significantly different from zero, and all of them are extremely low, indicating that performance on the topographical orientation test is not closely related to performance on the other tests given. It should be noted that the topographical orientation score is an error score, since it was the sum of the individual's deviations from the initial great circle heading for each city. (The odd-even reliability of this test was: Part I—.59; Part II—.65; and Total Score .75, when corrected by the Spearman-Brown formula, the reliability of the total score becomes .86). The test of topographical orientation showed the highest relationship to the Q score on the American Council on Education, Psychological Examination and the test of Mathe-

TABLE III.—Correlation between the Test of Topographical Orientation and Other Variables. (Note that the topographical orientation score is an error score.)

	Topographical Orientation			N
	Part I	Part II	Total Score	
1) Spatial Orientation (Guilford-Zimmerman, Part V)	.06	-.01	-.04	233
2) Spatial Visualization (Guilford-Zimmerman, Part VI)	-.13*	-.07	-.11	233
3) Aviation Classification Test	-.06	-.10	-.09	224
4) Mechanical Comprehension Test	-.14*	-.02	-.08	224
5) American Council on Education—Q	-.19†	-.19†	-.21†	213
6) American Council on Education—L	-.04	-.08	-.07	213
7) American Council on Education—Total	-.15*	-.17*	-.18†	233
8) Correctness of Expression	-.15*	-.14*	-.16*	233
9) Mathematics	-.18†	-.21†	-.21†	232
10) Personal Estimate of Orientation	-.19†	-.09	-.15*	233
11) Age	.01	.06	.04	231
12) Educational Level	.04	.08	.07	233

* r -.13 significant at .05 level.† r -.18 significant at .01 level.

matics. There is a less significant correlation between topographical orientation test scores and the cadets' own evaluation of their orientation and the test of correctness and effectiveness of expression.

No significant relationship was found between the test of topographical orientation and the test of spatial visualization, which is contrary to the results found by High (14), who reported an r of .51 for his group. The reason for the discrepancy between High's results and the results of this study may be found in the fact that (a) High's group contained only seventy-five individuals, and (b) High's group was much more heterogeneous, containing both men and women college students. It would appear, therefore, that in a cadet population, spatial visualization and spatial orientation are not related to topographical orientation, as measured by this test.

The data for Naval Aviation cadets confirm Woodring's findings that there is no relation between topographical orientation and age or intelligence; this is true notwithstanding the fact that Wood-

ring's population was highly heterogeneous (31, 32). As has been pointed out above, Woodring also found for his subjects (based on six geographical points) that the distribution of average scores was bimodal and the individuals could be classified into three fairly distinct groups. No evidence of bimodality was found for the individual cities as noted in the preceding section. Similarly, the total score for the topographical orientation test gives no evidence of bimodality; however, the scores do fall on a skewed distribution. For the group of two hundred and thirty-three persons used in this part of the study, the mean was 78.04 with a standard deviation of 33.40 and a range from 40 to 390. However, only four of these scores fell between 187 and 390, while only six scores fell between 138 and 179. If these ten scores are eliminated, the mean is slightly reduced to 73.19, while the standard deviation is greatly reduced to 16.80. The remaining scores show only slight skewness (index skewness equals .30) toward poor orientation as compared with a moderate index of skewness (.50) for the total population (12). The distribution of scores on the topographical orientation test can be described as being moderately skewed toward poor orientation, but there is no indication of bimodality in these data.

A composite personal orientation score was obtained for each of the subjects (items 5-12, Table IV) from the personal data listed in the procedure. Each of these eight items was assigned a weighted score depending upon which of the choices listed under that question was checked. The scores for the individual items were then summed and this total was assigned as a measure of the personal orientation. To determine whether there was a relationship between the answers to each of the questions and scores on the topographical orientation test, the following procedure was followed for each question: 1) the responses were sorted into like groups, i.e., all subjects who marked question 5, 'chiefly near home town', were placed in one group; 'chiefly in home state' in a second group, etc. 2) The scores on the topographical orientation tests for each of these groups were obtained and an analysis of variance was performed to determine where the differences between group performances were significant. 3) If a significant F was obtained, thus indicating that several groups differed significantly among themselves, a t test was run to determine those differences between means which were significant.

Questions 4, 5, 9, 11 and 12 provided F 's significant at the .01

TABLE IV. SUPPLEMENTARY DATA

- Name 1. Age (nearest year).....
- 2) Home city:, State:
- 3) Flight status: aviation training, ... stage;
..... pilot, ... nonpilot; approximate hours of flight
time
4) Encircle the highest educational level finished: High school—1, 2, 3, 4
years; College—1, 2, 3, 4 years; Post graduate.
- 5) Travel in the United States: chiefly near home town,
..... chiefly in home state, ... chiefly in home and nearby
states, ... one or two transcontinental trips (or trips over 1,500
miles), ... three or more transcontinental trips (or trips over
1,500 miles).
- 6) Travel outside the continental limits of the United States:
..... never, ... one or two visits to Canada or Mexico,
..... several trips to Canada or Mexico, ... one trip over-
seas, ... two or more trips overseas.
- 7) How many courses in geography or related subjects have you had in
the last three years?
- 8) Do you keep the direction of the major compass points (N, S, E, W) in
mind when riding as a passenger in a car?

Always	Very often but not al- ways	Sometimes if I have a reason	Occasionally when not familiar with the terrain	Almost never
--------	-----------------------------------	------------------------------------	---	--------------

- 9) Do you usually orient yourself in travel by means of some reference
point, such as a city or some other place whose location is familiar to
you; i.e., can you point to it if you wish?

Never	Occasionally if I have a special reason	Sometimes if possibility of becom- ing lost	Quite fre- quently	Always out of habit
-------	--	--	-----------------------	------------------------

- 10) If you are traveling in a strange area (city or country), do you try to
remember the number and direction of the turns you have made so
that you will be able to find your way back the same way if necessary?

Almost always	Frequently	Sometimes if I have a special rea- son	Occasionally if I know I will have to return the same way	Never
---------------	------------	---	--	-------

TABLE IV.—Continued

- 11) Do you keep a "mental map" of the area you are in, i.e., if you were required to do so could you reproduce a rough approximation of the area in which you find yourself as by drawing a crude map?

Never; I am not concerned about these things	Only if the area is well known to me	I do this as much as most people	Very often	Invariably through habit
--	--------------------------------------	----------------------------------	------------	--------------------------

- 12) Can you visualize a globe of the world with the continents and major bodies of land in their approximate positions with respect to each other?

Confident I have an accurate picture of globe	Believe I know positions of continents very well	My picture of globe is about average	Unsure of shape and position of some continents	I have a very poor picture of globe in mind
---	--	--------------------------------------	---	---

level and subsequently the differences between the means of the subgroups for each question were examined to determine those differences which were significant. For question 4, the group with a high school education was significantly better oriented ($P = .01$) than the group with one year of college. For question 5, the group which had answered, 'chiefly in home town', had poorer orientation ($P = .01$) than the groups which had answered the remaining four responses. For question 9, no significant differences were obtained between the means of the groups; this tends to indicate that although the samples were not drawn from a common population the difference is in variances and not in means. For question 11, the group which had answered, 'invariably through habit', showed better orientation ($P = .05$) than the groups which had answered, 'very often' and 'I do this as much as most people'. For question 12, the group which answered, 'My picture of the globe is about average', showed poorer orientation ($P = .01$) than the groups which answered, 'confident I have an accurate picture of globe' and 'believe I know position of continents very well'. Also, the cadets who answered, 'unsure of shape and position of some continents', showed poorer orientation ($P = .05$) than the cadets who answered, 'confident I have an adequate picture of globe'.

The choices for questions 8 through 12 were examined to deter

mine how closely the answers approximated the distribution of individuals normally expected to fall within the various classes of ratings (12); assuming that the trait that is rated has a normal distribution in a fairly large unselected group. None of the questions met the criterion. It must be pointed out that this lack of continuity of the items of these questions no doubt had an important effect upon the variable results which were obtained. It is recognized that this weakness must be remedied in any future use of a personal orientation questionnaire. For this reason, a minimum amount of emphasis is placed upon the results previously mentioned. However, it is also noted that the results obtained point favorably to the feasibility and utility of such a revised personal biographical instrument to augment any future studies on topographical orientation.

Discussion—The data show that the factors measured by the topographical orientation test are unique. Very low correlations were found between these scores and those on two other orientation tests, two intelligence tests and a mechanical aptitude test. A similar result was found for a variety of personal data, though some individual items of the personal questionnaire did contain items which were significantly related to topographical orientation. The fact that those who indicate travel, 'chiefly near home town', show poorer topographical orientation than those with more extensive travel experience might have been anticipated. However, it is interesting to note that the other categories show no significant differences. The significantly better orientation of individuals with no more than high-school education as compared with those with college training tends to support Woodring's (32) finding that persons with lesser amounts of schooling do better on such tests. The non-college cadets come from a group which has had a relatively high amount of travel experience in the Navy which is undoubtedly an important factor. It is apparent that the current selection devices in the Navy do not include a variable related to topographical orientation as measured by this test. The relation of this composite score to success in Naval Aviation Training is not known. However, a study currently underway will obtain data on this point.

It will be noted that all of the deviations measured by this test were measured from initial great circle headings which would give the correct initial heading to reach the city by the shortest possible

route. Similar calculations were made using the mean heading indicated by the group. This attenuated the error scores, but it did not change the intercorrelations significantly. Scatter plots of the data also indicated that the elimination of the cadets having the most widely deviant scores would not cause any significant change in the correlations.

SUMMARY

A test of topographical orientation and seven other tests were given to two hundred and forty-two Naval Aviation cadets to study the types of errors in orientation they make, to identify the factors which contribute to this type of disorientation, and to determine what relation there is between topographical orientation and certain other psychological factors. In the test of topographical orientation, large errors were found to be the rule for orientation to individual geographical points. Deviations of the group from proper orientation were as great as 65.1° , while some individual deviations from the proper orientation approximated 180° , the maximum possible error. Errors for cities located abroad were greater than for cities within the United States. Deviations from Mercator headings were large, but these errors were much less than for deviations from great circle headings, particularly for cities outside of the continental limits of the United States. Correlations between total score on the test of topographical orientation and other psychological tests were .21 or below; only three of them being significantly different from zero. Naval Aviation cadets are poorly oriented to geographical points, particularly as they actually occur on the globe. The notion of great circle relationships between distant cities does not appear to enter significantly into their judgments of the geographical points. Topographical orientation appears to be a unique characteristic which is independent of the factors measured by the current selection battery for cadets as well as certain other factors such as mathematics, spatial orientation, and spatial visualization.

REFERENCES

- 1) A. A. Ageton. *Dead reckoning, Altitude and Azimuth Table*. Washington, D. C.: U. S. Navy Hydrographic Office, No. 211, U. S. Government Printing Office, 1940.
- 2) A. Binet. "Reverse illusions of orientation." *Psychol. Rev.*, 1894, 1, 337-350.

- 3) H. R. De Silva. "A case of a boy possessing an automatic directional orientation." *Science*, 1931, 73, 393-394.
- 4) F. N. Freeman. *Psychology of the Common Branches*. Boston: Houghton-Mifflin Co., 1916. (Chapter VIII)
- 5) J. J. Gibson. *The Perception of the Visual World*. Boston: Houghton-Mifflin Co., 1950.
- 6) J. P. Goode. *School Atlas*. New York: Rand McNally and Co., 1943.
- 7) T. Gordon. *The airline pilot: a survey of the critical requirements of his job and of pilot evaluation and selection procedures*. Washington, D. C.: C.A.A. Div. of Research, Report No. 73, 1947.
- 8) A. Graybiel. "Spatial disorientation in flight." U. S. Naval School of Aviation Medicine, Research Report No. NM 001 059.01.24. (also in *The Military Surgeon*, 1951, 108, 287-294).
- 9) D. R. Griffin. "The sensory basis of bird navigation." *Quart. Rev. Biol.*, 1944, 19, 15-31.
- 10) — "Topographical orientation." (Ch. 17 in C. G. Boring, et al, *Foundations of Psychology*.) New York: Wiley, 1948.
- 11) D. R. Griffin and R. J. Hock. "Experiments on bird navigation." *Science*, 1948, 107, 347-349.
- 12) J. P. Guilford. *Psychometric Methods*. New York: McGraw-Hill, 1936.
- 13) J. P. Guilford and W. S. Zimmerman. *The Guilford-Zimmerman Aptitude Survey*. Beverly Hills, California: Sheridan Supply Company, 1947.
- 14) W. S. High. *An Investigation of Geographical Orientation and Its Relation to Spatial Visualization*. Unpublished M.A. thesis, San Jose State College, San Jose, Calif., 1951.
- 15) G. F. Howe. "The teaching of directions in space." *J. of Geog.*, 1931, 30, 298-304.
- 16) — "The teaching of directions in space." *J. of Geog.*, 1932, 31, 207-210.
- 17) J. I. Lacey and S. W. Niehaus. "Orientation tests" in J. P. Guilford *Printed Classification Tests, Report No. 5*, Army Air Forces Aviation Psychology Program Research Reports. Washington, D. C.: U. S. Government Printing Office, 1947.
- 18) F. E. Lord. "A study of spatial orientation of children." *J. ed. Res.*, 1941, 34, No. 7, 481-505.
- 19) E. Meyer. "Comprehension of spatial relations in pre-school children." *J. genet. Psychol.*, 1940, 57, 119-151.
- 20) W. B. Michael, W. S. Zimmerman, and J. P. Guilford. "An investigation of two hypotheses regarding the nature of the spatial relations and visualization factors." *Ed. and Psychol. Measurement*, 1950, 10, 187-213.
- 21) K. C. Montgomery and E. G. Heinemann. "Concerning the ability of homing-pigeons to discriminate patterns of polarized light." *Science*, 1952, 116, 454-456.
- 22) H. E. Page. "Vertigo incidence among naval aviators." U. S. Naval School of Aviation Medicine, Research Report No. NM 001 063.01.16, 15 1950.

- 23) D. Ridgley. "The teaching of directions in space and on maps." *J. of Geog.*, 1922, 21, 66-72.
- 24) W. Rowan. "Homing, migration and instinct." *Science*, 1945, 102, 110-111.
- 25) T. A. Ryan and M. S. Ryan. "Geographical orientation." *Amer. J. Psychol.*, 1940, 53, 204-215.
- 26) L. F. Shaffer, et al. "Surveys of experiences of returned personnel," in F. Wickert, Ed., *Psychological Research Problems of Redistribution*, Army Air Forces Aviation Psychology Program, Research Report No. 14. Washington, D. C.: U. S. Government Printing Office, 1947.
- 27) C. C. Trowbridge. "Fundamental methods of orientation and imaginary maps." *Science*, 1913, 38, 97.
- 28) W. E. Vinacke. "Illusions experienced by aircraft pilots while flying." *J. Aviat. Med.*, 1947, 18, 308-325.
- 29) H. C. Warren. "Magnetic sense of direction." *Psychol. Bull.*, 1908, 5, 376-377.
- 30) H. A. Witkin. "Studies in geographic orientation." *Yearbook Amer. Phil. Soc.*, 1946, 152-155.
- 31) P. D. Woodring. "A technique for the investigation of direction orientation in human beings." *Michigan Acad. Arts and Sciences*, 1938, 24, Pt. iv, 147-152.
- 32) —. "An Investigation of Direction Orientation in Human Adults." Ab. Doctoral Dissertations, No. 29. Columbus: Ohio State Univ. Press, 1939.

STUDENT DIFFERENCES IN COMPETENCY IN EDUCATIONAL PSYCHOLOGY PRIOR TO FORMAL TRAINING¹

WILLIAM W. LYNCH, JR.

Indiana University

The research to be reported here is an attempt to analyze and study two types of competence that are commonly stressed in the training of prospective teachers in educational psychology: (1) familiarity with the principles, facts, and value judgments most typically presented in textbooks in educational psychology, and (2) ability to interpret information about an individual child's behavior problems, as presented in written case-study form, and make valid recommendations for the solution of these problems. The need for such analysis can be supported by a review of the literature of the past several decades dealing with the content and teaching of educational psychology. In recent years, particularly, the discussions of ways and means of improving instruction in educational psychology have stressed the opinion that the psychological competencies needed by teachers are not fulfilled by familiarity with principles, facts, and authoritative judgments, and that this familiarity and its applicative concomitants are inextricably involved with the prospective teacher's total personality make-up. Just what the nature of these relationships may be is far from clear. It was the intent of the present study to probe this complex of relationships by studying the two abilities mentioned above as measured in a population of teachers college students planning to become elementary teachers who had received no formal instruction in psychology. By selecting such an untrained group it was possible to discern the extent of differences in these two abilities prior to training and to study the relationship of these differences to several psychological characteristics and background experiences of a primarily non-academic nature. The related factors selected were: tested intelligence; personal and social adjustment; occupational interests; attitudes toward child care and discipline; and amount of past experience in caring for, teaching, or directing children.

¹Based in part upon the writer's dissertation submitted for the Ph.D. at Yale University, 1950.

HYPOTHESES TO BE TESTED

On the basis of a review of experimental studies and published authoritative opinion, and a logical analysis of the factors involved in the study, the following seven hypotheses were formulated for testing:

I) Students who have had no instruction in educational psychology differ significantly in their familiarity with the facts and principles of this field that are contained in textbooks. Some of the facts and principles are familiar to a majority of these students.

II) Students who are most familiar with the content of educational psychology are best able to interpret case-study data and to select valid recommendations for remedial measures.

III) Among persons with no formal training in psychology there is a positive relationship between tested intelligence and both (1) familiarity with facts and principles and (2) ability to interpret case study data correctly.

IV) Prospective teachers with no formal psychological training who have had extensive experience with young people are better informed in the content of educational psychology and are more competent to interpret case material.

V) Prospective teachers prior to psychological training who have measured interests in social activities and in the care and direction of others are more proficient in both knowledge and applicative ability than those who show little or no interest in such activities.

VI) Good social and personal adjustment is positively correlated with the informational and applicative criteria of proficiency in educational psychology.

VII) Persons who manifest strict authoritarian attitudes toward child behavior are poorly informed in educational psychology and are relatively lacking in the ability to interpret case material.

SUBJECTS

Four hundred and forty freshmen enrolled in three Connecticut state teachers colleges that prepare elementary teachers were selected to test the hypotheses. These were divided into subgroups. The numbers in each subgroup and the hypotheses to be tested are shown in Table I. All testing was done during the first half of the fall semester of the 1949-1950 academic year.

TABLE 1. ARRANGEMENT OF GROUPS FOR TESTING

College	N	Tests	Hypotheses
A	220	Subtest on 'Personality'	I, II, III
	47	Subtest on 'Growth and Development'	
	45	Subtest on 'Individual Differences'	
	48	Subtest on 'Motivation'	
	48	Subtest on 'Learning'	
	57	Subtest on 'Problem-solving'	
	58	Subtest on 'Evaluation'	
	69	The Case of Mickey Murphy	
B	131	Abbreviated test of Information	II, III, V, VI
		The Case of Mickey Murphy	
		The Minnesota Personality Scale	
		The Kuder Preference Record	
C	93	Abbreviated Test of Information	II, III, IV, VII
		The Case of Mickey Murphy	
		Experience Rating Scale	
		Scale of Attitudes Toward Child Behavior	

CRITERION TESTS

1) *Content Tests*.—A comprehensive battery of 432 objective test items was constructed by the writer and an assistant. These items were based upon the content of fourteen textbooks in educational psychology published between December, 1945, and August, 1949. The battery was divided into seven subtests covering the following topics most commonly treated in the textbooks:

Individual differences, growth and development, personality, motivation, learning, problem-solving, and the evaluation of school achievement. The curricular validity of the battery was checked by submitting the items to four educational psychologists whose training and teaching experience presumably qualified them to judge the validity of the items. The split-half reliability of the seven subtests, corrected by the Spearman-Brown formula, ranged from .21 to .67, the median being .53. An abbreviated form of the entire battery was constructed by selecting items at random from the original subtests in proportion to the amount of space devoted to the subtest topics in the fourteen source textbooks. This

form had a reliability coefficient of .58, as corrected by the Spearman-Brown formula.

2) *Case-Study Test*.—A published case-study test, *The Case of Mickey Murphy*², was selected as the criterion measure of ability to diagnose a problem involving a young person in a school situation and to make valid recommendations for remedial procedures. This test was based upon actual case-study data and was validated by the judgments of experts. The odd-even reliability obtained on the group tested in this study was .64 when corrected by the Spearman-Brown formula.

DEPENDENT VARIABLE MEASURES

1) *Intelligence*.—Scores on the *California Test of Mental Maturity, Advanced '47 S-Form*, were obtained for all subjects. This test had been administered as an admissions test prior to this investigation.

2) *Experience with Children*.—A check list of activities such as baby-sitting, Sunday-school teaching, etc., was devised and accompanied by a rating scale designed to indicate the extent of experience in the care and direction of children. These instruments were based upon a preliminary questionnaire study of the experiences of a group of education and psychology students. The freshmen in this study were asked to check all of their experiences involving the care and direction of young people, and then were asked to estimate the extent of this experience when compared with that of full-time classroom teaching in terms of actual time spent in contact with children. Length-of-experience intervals ranged from 'one to three months' to 'two or more years'.

3) *Interests*.—*The Kuder Preference Record—Occupational* was used as a measure of occupational interests. The data from male and female subjects were treated separately.

4) *Personal and Social Adjustment*.—*The Minnesota Personality Scale* was administered to the freshmen used in testing Hypothesis VI. Male and female data were treated separately.

5) *Attitudes toward Child Behavior*.—The writer constructed a nineteen-item attitude scale based upon the written opinions of students in classes in psychology and child development on a child behavior problems such as discipline, freedom of choice

² Warren Baller, *The Case of Mickey Murphy*, Lincoln: The University of Nebraska, 1943.

for children in everyday matters, the use of punishment and reward, academic standards, and the characteristics of a successful public-school teacher. The scale was designed to measure on an 'authoritarian-permissive' continuum. The Likert technique for construction and scoring was used. No direct check on validity could be made, but a preliminary scale was submitted to three experienced instructors in education and psychology who were asked to indicate the responses to each item that would reveal in their opinion attitudes at each end of the postulated continuum. Items on which the judges disagreed or which were judged to be either ambiguous or not related to the postulated continuum were discarded. The split-half reliability, corrected by Spearman-Brown formula, was .58. The test-retest reliability, with an intervening interval between testings of about four months, was .65 for 34 subjects.

PROCEDURE

Hypothesis I.—The schedules of the two hundred and twenty freshmen subjects in college A precluded the possibility of administering the entire battery of content tests to each freshman. Therefore, the group was divided into five sections and three of these sections were administered three content tests each, and two sections were administered one content test and the case-study test (see Hypothesis III). All groups took one of the subtests in common so that a rough check could be made of the equivalence of the five freshman sections (Table I).

The content test scores were treated both as raw scores and as 'true per cent scores' which were based upon theoretical 'zero' scores (scores theoretically obtainable by chance). The mean 'true per cent scores' and the ranges of raw scores were compared for the seven subtests to determine whether a distinct profile of subtest scores existed among freshmen. The freshman scores were then compared with the distributions of scores obtained from seniors in the same college and the extent of overlap was determined. Finally, an item analysis for all content tests was carried out in order to determine significant differences in responses of freshmen and seniors.

Hypothesis II.—One hundred and thirty-one freshmen in college B and ninety-three freshmen in college C were tested with the 'deviated' content test which was derived from the total

content test battery. The same groups were also administered the case-study test. Pearson product-moment correlation coefficients between content scores and case-study test scores were computed and corrected for attenuation. The discovery of significant correlations between IQ and the two criterion tests made it desirable to compute first order partial r 's between the two criterion scores partialing out IQ.

Hypotheses III, V, VI, and VII.—These hypotheses were tested by computing zero-order correlations between each of the two criterion measures and the measures of intelligence, occupational interest, personality, and attitude. In each case the abbreviated content test was used.

Hypothesis IV.—Hypothesis IV was tested by determining the significance of the difference between the mean criterion scores obtained by the top and bottom twenty-five per cent of the group as determined by self ratings of length of experience.³

A second method of testing Hypothesis IV was devised by re-sorting the self-ratings together with the check list data and selecting two groups of high and low experience as determined by the combined criteria of self-ratings and the writer's judgment of the check list data.

RESULTS

1) Hypothesis I, positing the existence of differences in familiarity with the content of educational psychology among beginning prospective teachers, appears to be substantiated by the data obtained. In the first place, on both the separate topical area tests and on the total test the freshmen scored significantly above the theoretical 'zero' score obtainable by chance. In the second place, these freshman scores generally ranged from a low around the 'zero' score to a point sufficiently high to result in a mean overlap of 7.1 per cent of freshman scores above the median scores of seniors who had completed at least three courses in psychology (general, child, and educational psychology) in addition to methods courses and practice-teaching. These two general findings indicate that freshmen do differ significantly in their knowledge of the facts and principles in educational psychology. These findings are summarized in Table II.

³ The small number of experience categories and a skewed distribution prevented a correlation treatment.

TABLE 2.—SUMMARY OF RESULTS ON CRITERION TEST OF EDUCATIONAL PSYCHOLOGY CONTENT

Group	N	Test	Theoretical Chance Score	Median	Range	Median Senior Score*
College A	47	Growth and development	24.1	33.6	17-46	38.7
	45	Individual differences	24.5	31.2	25-41	38.2
	47	Motivation	25.5	37.4	24-45	43.8
	48	Learning	24.4	35.5	29-48	39.9
	57	Problem-solving	23.0	34.6	22-43	40.5
	58	Evaluation of school Achievement	22.2	32.6	20-40	38.8
All Sections	220	Personality	22.7	33.2	20-45	41.9
Colleges B and C	224	Abbreviated form	22.0	33.0	20-45	38.9

* Scores obtained on 112 Seniors in College A, 52 Seniors in Colleges B and C.

The freshman responses to the separate items in the content battery, when compared with the senior responses, indicate general familiarity with certain facts and principles in educational psychology, at least at a recognition level. Of the four hundred and thirty-two items in the total battery, two hundred and eight were answered correctly by a significant majority of freshman subjects. Two hundred and fifty items showed no significant difference between the freshman and senior percentage answering correctly. Of the one hundred and eighty-two items on which there were significant differences between the two groups, twelve were answered correctly by a significantly larger proportion of freshmen than seniors.

It was difficult to categorize the items which most freshmen answered correctly. Generally speaking, they consisted of generalizations expressed in nontechnical language. In some cases freshman superiority over senior responses plainly indicated that general information or 'common knowledge' that is valid psychological fact had been distorted by seniors through poor comprehension of psychological data during training.⁴ The mean freshman scores

⁴ For example, ninety-six per cent of freshmen recognized that 'drill facilitates learning,' while only seventy-six per cent of the seniors answered

on the subtests, while showing some small differences that approached statistical significance, do not warrant a clearcut conclusion as to the relative familiarity with the seven topical areas covered by the content test battery.

A word of caution is in order in interpreting the results pertaining to Hypothesis I. No attempt was made to discern evidence of response sets which conceivably may have affected some of the item-analysis results. Secondly, it is not redundant to point out that the behavior of these subjects being measured was most probably a very low order recognition type of response and the term familiarity, used in the statement of the hypothesis, should be interpreted accordingly.

2) Hypothesis II, stating that there would be a positive relationship between familiarity with textbook content and ability to interpret case-study data, was definitely confirmed. For College B and College C freshmen the uncorrected correlations between the content test scores and the case-study test scores were .46 and .37, respectively, (both significant at the .01 level). When corrected for attenuation these coefficients were .76 and .61. Partial r 's were computed, holding tested intelligence constant. The partial r 's were .64 and .55, respectively. These results indicate that theoretically the relationship between the two variables is fairly high; somewhat higher, in fact, than those found in most investigations of the relationship between knowledge and ability to apply.⁵

3) The positive relationship between tested intelligence and the two criterion measures of psychological proficiency posited in Hypothesis III was borne out. College B and College C freshmen yielded correlations of .37 ($SE_r = .08$) and .43 ($SE_r = .06$) for the content test scores, and .39 ($SE_r = .08$) and .35 ($SE_r = .09$) for the case-study test scores. Correction for attenuation raises each of these from ten to fifteen points.

this correctly. Obviously there was confusion between the effect of drill upon certain learning outcomes with the undesirability of drill when applied to certain educational objectives.

⁵Maurice B. Troyer and Horrocks, John B. "Case-study tests of ability to use knowledge of human growth and development," *Educational and Psychological Measurement*, 7: No. 1, 23-26, 1947. Ralph W. Tyler, "The relation between recall and higher mental processes," Chapter II in Charles H. Judd, *Education as Cultivation of the Higher Mental Processes*. New York: Macmillan Company, 1936, pp. 6-17.

4) Two methods of selecting high- and low-experience groups from the freshmen of College C and comparing these groups' mean scores on both criterion tests yielded no differences that were statistically significant.

5) Interest test scores in 'social service' for male and female subjects yielded correlations with the criterion tests that did not differ significantly from zero.

6) Separate correlation of five subtest scores on the *Minnesota Personality Scale* with criterion test scores for male and female groups did not differ significantly from zero.

7) Hypothesis VII, which assumes a relationship between attitudes toward child behavior on an 'authoritarian-permissive' continuum, seemed to be confirmed. The correlation of raw scores on the attitude scale with the criterion tests were: content test, .24 ($SE_r = .10$) (significant at the .05 level but not at the .01 level); case-study test, .32 ($SE_r = .10$) (significant at the .01 level). These indicate a tendency for those with 'permissive' attitudes to score higher on both tests. When corrected for attenuation these correlations become .51 and .52, respectively.

These results are not surprising when an examination is made of the separate items on both tests. Recognition of the 'correct' answers on the content test frequently requires the testee to make a value judgment rather than a decision of correctness of fact or theory. The sampling of the textbook content required for this test included judgments of textbook writers concerning desirable educational practice. The same point may be made with regard to the case-study test. Here also many questions are concerned with remedial practices in a school situation. A person generally in favor of more lenient or permissive handling of educational problems would be apt to do better on either test.

SUMMARY AND CONCLUSIONS

Freshmen in three Connecticut state teachers colleges were administered tests of information based on the content of fourteen textbooks in educational psychology and a case-study test purporting to measure ability to interpret and make valid remedial recommendations concerning an educational problem of a psychological nature. Scores on these tests were examined to determine: (1) the extent of differences in ability to recognize correctly facts, principles, and authoritative opinion as contained in educational

psychology textbooks prior to any formal training in this subject; (2) the relationship between familiarity with textbook content and ability to interpret case-study data; and (3) the relationship of the two measures of psychological proficiency to measures of intelligence, experience in care and direction of young people, interest in social service, social and personal adjustment, and attitudes toward child behavior and discipline.

It was found that freshmen demonstrate some familiarity with the facts and generalizations of educational psychology prior to training and that there is a significant spread of differences in such familiarity. Content familiarity and ability to interpret case-study data are positively related to a relatively high degree.

Tested intelligence and attitudes toward child behavior are significantly related both to the knowledge and to the applicative aspects of educational psychology. Experience with children, vocational interest in social service, and personal and social adjustment as measured by the paper-and-pencil instruments used, show no significant relationship with the two criteria of proficiency.

It should be emphasized that the study was an exploratory one, using instruments whose validity and reliability were suitable for a general, theoretical examination of the attributes concerned and not for the precise testing of exact hypotheses. In spite of this general nature of the study, several implications pertaining to educational psychology in teacher education are suggested:

1) The study's confirmation of the existence of considerable heterogeneity of psychological proficiency among prospective teachers prior to formal training in educational psychology reinforces the general principle that instruction in this field, as in others, needs to be geared to the differences that exist. Furthermore, not only does this point to the need for caution in the interpretation of objective tests used as measures of achievement in educational psychology courses, but it also points to the need for more penetrating analyses of the psychological proficiencies of teachers-in-training both before and after instruction in educational psychology. Until such study is made it would seem difficult to set up standards of psychological proficiency for teachers that are realistic and valid.

2) The finding of a positive relationship between the attitude measure and the criterion tests suggests a host of problems concerning the rôle of values in psychology instruction. The fa 19 t

close examination of both criterion tests revealed that the testee was required to make judgments not only of correctness of fact and general principles that account for human behaviors, but also of the correctness of certain procedures that depend upon value choices as well, suggests that educational psychologists need to attend more to the ways in which these value positions are learned. The question of when and how the 'should' questions are best handled in relation to the questions of 'what' and 'how' are in need of study.

3) The apparent lack of relationship between the amount of experience with children and the two criterion measures reinforces the conclusion that first-hand experience *per se* is not likely to enhance one's psychological understanding. Current emphasis in some teacher-education programs upon early field experiences may need to take into account more critically the types of guidance needed in such experiences if mature and valid understandings are to result.

BOOK REVIEWS

B. F. SKINNER, *Science and Human Behavior*. New York: The Macmillan Company, 1953, pp. 461. \$4.00

Psychologists should have known what Skinner stands for from familiarity with his two previous books, *The Behavior of Organisms* and *Walden Two*, the former being a report of his investigations into the psychology of learning and the second being a novel which deals with the problem of Utopia. However, Skinner has really revealed his hand in this latest book, and has proclaimed himself to be a 'neo-behaviorist', taking on the mantle which John B. Watson let lapse over two decades ago. Skinner says in brief that behavior is the only datum concerning the mind with which science can deal and so he conceives of psychology strictly as the science of behavior. Skinner believes that the principal mission of science is to discover relationships that have a certain degree of generality and permanence. He feels that the only relationships with regard to behavior with which science can deal are those that connect behavior with the outer physical environment. Therefore, to Skinner behavior is the dependent variable and the environment is the independent variable, and the task of scientific psychology is to find the functional relationships between these two sets of variables.

Skinner has no use for so-called intervening variables—constructs which other psychologists have found necessary to set up in order to explain mental processes. A large part of this book is concerned with the endeavor to explain away the many intervening variables which have been proposed both in the common-sense world of everyday life and also by professional psychologists, and to reinterpret the phenomena which these intervening variables represent in terms of situation and response. Sometimes it is almost laughable to find Skinner taking an everyday concept which is accepted without difficulty by the common man and retranslate it into his own terms of reinforcement, extinction, discrimination and induction. What Skinner has done in the book is to set up a new psychology which will be consistent throughout in terms of a few principles which he believes to be basic to a psychology which is scientific.

But this book is more than a mere psychology, for in addition it has a program and a philosophy. Skinner says that "science s

more than a mere description of events. It is an attempt to discover order, to show that events stand in lawful relation to other events." (p.6) "Science not only describes, it predicts . . . nor is prediction the last word. To the extent that relevant conditions can be altered or otherwise controlled, the future can be controlled." This then is the keynote of the book. Skinner is interested in building a science of behavior in order to be able to control behavior.

After he finishes making his analysis of behavior, he turns to the problem of control. First, in Chapter 15, he describes self-control, and later, in Chapter 20, what he calls personal control and, in Chapter 21, group control. This is followed by a discussion of various controlling agencies in society, including government and law, religion, psychotherapy, economic control and education.

The final section of the book discusses the problem of human control in general, including such issues as, who shall control and for what purposes. Skinner belittles value judgments and believes that they have no place in the science of behavior. He then proceeds to propose his own basic value and make it the fundamental guiding principle on which all decisions with regard to control shall be made. In doing so he draws a parallel between the process of natural selection in the biological world and the process of selection of behavior in society. The guiding principle in biological selection is the survival of the species. In similar fashion he believes that the guiding principle in the selection of behavior is the survival of a culture. He says "a given culture is in short an experiment in behavior" (p 430) and he believes that the final criterion of behavior is the survival of a culture, and the problem of estimating the survival value of a culture is one for experimental determination. Skinner places greater value on the culture than on the individual, with all the social implications that such a point of view carries with it. He recognizes that this emphasis on the culture goes contrary to the emphasis in western democracy on the individual, but he believes that his thinking leads him inevitably to this conclusion.

Section II of the book is a summary of Skinner's views about learning, and anyone who wishes a compact statement of Skinner's point of view will find it in this section. Skinner, of course, was one of the first to propose a two-factor theory of learning, and this discussion is repeated in this book in terms of stimulus substitution

and response substitution. Skinner, however, does not take the extreme point of view which Mowrer has adopted in believing that these two kinds of learning are mutually exclusive. If this reviewer interprets Skinner correctly, he believes that most learning involves both stimulus substitution and response substitution. Skinner gives full recognition to the principle of secondary reinforcement which he calls 'conditioned reinforcement'.

He has a very interesting discussion of intermittent reinforcement and shows some of its practical implications. It will be remembered that some of the first experimental findings with regard to intermittent or partial reinforcement have come out of Skinner's researches. Skinner pays much attention to aversive behavior and avoidance behavior and there is a good discussion of punishment, drawing on contributions of Estes who did his researches under Skinner's supervision.

Skinner strives to be consistent throughout the book in avoiding the intervening variable, but when he comes to what he calls 'private events in a natural science' and the topic of the self, he occasionally slips from the rigorous position which he has attempted to adopt. Although he apparently is reporting Freud's point of view on page 284, he says "multiple selves or personalities are often said to be systematically related to each other. Freud conceived of the ego, superego, and id as distinguishable agents within the organism . . . the superego and the id were inevitably opposed to each other and Freud conceived of them as often in violent conflict. He appealed to a third agent—the ego—which, besides attempting to reach a compromise between the id and the superego, also dealt with the practical exigencies of the environment. We may quarrel with any analysis which appeals to a self or a personality as an inner determiner or action, but the facts which have been represented with such devices cannot be ignored. The three selves or personalities in the Freudian scheme represent important characteristics of behavior in a social milieu."

Apparently Skinner has made no effort to translate these intervening variables—the id, the superego and the ego—into the terms of his behavioristic categories but says "the facts which have been represented with such devices cannot be ignored."

This reviewer read with interest Skinner's discussion of the five controlling agencies in society, but feels that he has unfairly depicted at least two of them; namely, religion and education. 19

ner has described institutionalized religion and education as they now exist in current society, and has not given sufficient credit to religion and education which may be more of an inner, personal matter. In neither has he given sufficient credit to self-control, but has thought of both of these agencies as representing a fully externalized control by priests and teachers. As a matter of fact, because Skinner is so concerned with only relating behavior to external events he fails to give proper place to so-called inner or self-controls. In the discussion of self-control he rightly points out that all self-control is an outgrowth of external control. He fails, however, to give recognition to the fact that these external controls may have existed in the past and dropped out of the present picture so that for all intents and purposes these controls have been internalized or introjected. In omitting the recognition of these internalized controls as factors in religion and education he has given only a narrow and partial presentation of these controlling agencies.

Skinner seems to believe that control in education is very much a matter of punishment, and apparently does not recognize other types of control which are of a more positive nature. Evidence which this reviewer has gathered points conclusively to the fact that those teachers who depend most upon punishment as a means of securing control are among the least successful as teachers. This would point to the fact that there are other types of control in education to which Skinner has not given his attention.

The publishers of this book in their advertising refer to this as "an introductory psychology book." Certainly Skinner himself can be under no illusions that this is an introductory text. As a matter of fact, even the advanced student may find this book tough reading because it presupposes a rather thorough acquaintance with a vast amount of experimental material. This is a book without a statistical table and without the formal report of a single experiment. The unsophisticated reader may get the impression that Skinner has derived much of his material from the observation of and experimentation with pigeons because there are so many references to Skinner's later work with pigeons.

Any serious student of psychology, however, who wants to learn exactly what Skinner's system is and what applications he would

make of it to the practical world of affairs, will find this book a helpful reference for reading and study.

PERCIVAL M. SYMONDS

Teachers College, Columbia University

LESLIE PHILLIPS AND JOSEPH G. SMITH. *Rorschach Interpretation: Advanced Technique*. New York: Grune & Stratton, 1953, pp. 385.

Principles and procedures for Rorschach interpretation are nowhere more carefully explained than in this book. The authors set forth in detail many procedures clinical interpreters have otherwise had to pick up by observation and imitation. The procedure is not thereby oversimplified or stereotyped. The authors firmly believe that every fact must be related to every other, and they carry their suggestions to the utmost frontier of Rorschach interpretation. No cue is too faint, no relationship too speculative, to be included as a suggestion.

A unique feature of the book is that the authors are aware that their interpretations may be invalid. Their first paragraph is refreshing: "This book is intended primarily as a practical clinical reference. It is composed largely of statements about relationships between Rorschach performance and other behavior. Unfortunately, few of the relationships which are asserted have been corroborated; most represent guessed-at laws. They are perhaps best employed with equal parts of faith and skepticism." The writers warn that one proceeds by successive refinements, setting up tentative inferences on Card I which constitute predictions about subsequent cards; they can therefore be confirmed or modified (p. 261). Of course the hypothesis that results after ten cards then requires its "cross validation" from further data. It is in not stressing this need for testing the Rorschach interpretation that writers on the test err. Despite their sophistication and their excellent section on errors, Phillips and Smith do not succeed in maintaining their rigorous or skeptical attitude throughout the book.

Wherever the writers do reexamine current doctrine, they make contributions. They often explain why interpretations are proposed. They develop much improved criteria for scoring the junction
19

and form level. They describe the interpretation of 'style' or non-scored behavior most helpfully. They present new developmental norms derived from theses by Hemmendinger, Siegel, and Friedman. The reviewer does not feel that their genetic approach to interpretation is always helpful. To say that an adult has a score common among children is not explained by describing this as a "genetically early level," or describing him as fixated at or regressed to this level. The significant point is that he has now acquired a positive way of behaving with his present organic potential, presumably because this way of behaving is relatively rewarding. Wherever analysis in genetic terms appears, it seems to crowd out discussion of learning processes or dynamics generally.

Recognizing the good features of the book, one still is startled by much of its content. "Shock on Card I is evidence for an unresolved and intense relation with a mother figure . . . Shock on Card II is characteristic of . . . uncontrolled destructive impulses." Remarks like "any time limits are characteristic of persons with obsessive-compulsive character features," "The concept 'teeth' . . . appears to be associated with masturbation and/or sibling rivalry as a reaction to frustrated dependency needs." Such comments are generally expanded in a way which gives them plausibility, yet the fact that the cables of evidence anchoring them to reality have all the tensile strength of moonbeams, and the authors' introductory warning that what they say may be either true or false, cause one to question just how such writings can have a place in a science.

If this book were read by a man from some 'ivory tower' who happened not to know that the Rorschach is currently taken seriously, it would sound to him like *Ralph 1245C+* sounded to its first readers. Good science fiction is written with logic, any known facts, and imagination; at its best, its plausibilities foreshadow some things that science later establishes. Perhaps it is fairer to say that Rorschach interpretations are to the science of psychology as historical novels are to history. The writer seeks internal consistency and plausibility, with due regard for established fact. Where there is a gap in fact, the scholarly historian acknowledges the gap, but the fiction writer fills it with what might reasonably have happened, i.e., with an artistic creation.

The argument between scientist and clinician disappears if we see that clinical interpretation is an art at the present time. In the early days of testing it was a technology, wrong at times but never

seeking to be creative or intuitive. Now, instead, the clinician is first of all a portraitist who abhors unfilled areas on the canvas. This occurs because there is a practical demand for answers about people, not for uncertainties. Society has rewarded this new species of artist; we can scarcely criticize him for satisfying the demand. New entrants in the field want to be successful artists, and books like the present one which tell how to complete an attractive and convincing portrait are wanted. When authors like Phillips and Smith can also help keep portraits closer to reality than they would otherwise be, that is so much gained. Once we grant that the clinician is going to provide descriptions which go beyond proven fact, we should welcome any book which contributes to making him a good non-scientist.

The scientist cannot complain if the portrait is used as a guide to further inquiry about the patient or about personality in general. Both science fiction and historical fiction have suggested leads to serious investigations of their fields. This emphatically demands that the clinician's suggested interpretations should send further study in promising directions, not up 'blind alleys'. There is good reason to think that the Rorschach or other clinical interviewing methods can do this often enough to be useful, even though, according to the evidence, their suggestions on any case are not correct enough to be trusted without further support.

The scientist should object when he is told that he or the practitioner who receives a Rorschach analysis should believe it. Among all the books yet written on the Rorschach, the book under review comes closest to filling a practical demand for training clinicians to be artistic portraitists, without making unacceptable claims that such portraits are dependable.

LEE J. CRONBACH

University of Illinois

ERNEST GARDNER. *Fundamentals of Neurology*. Philadelphia: W. B. Saunders Company, 1952, pp. 359.

Neurology is, in this second edition of the textbook on the subject as in the first edition, presented as a series of basic concepts—revised in the light of growing knowledge in the field. The material selected for inclusion is based upon dissection and mice jumpic studies of the nervous system, experimental data, and their 19— of

neurological disorders. In the nineteen chapters of this small volume are included a consideration of functional as well as structural topics. Also included is a glossary of the many new terms of Greek and Latin sources, and an index. Each chapter ends with a brief biographical sketch of the people who have contributed to the topics considered in the chapter.

Psychologists will be interested in this text on neurology as such and also in the treatment of the problems of learning and language mechanisms in the last chapter called "Associative Functions of the Cerebral Cortex". Treatment of learning and conditioned responses are presented in a very compact manner, and the frame of reference for the consideration can be gathered from the psychological sources used by the author. These include Hilgard and Marquis' *Conditioning and Learning*, Lashley's *Brain Mechanisms and Intelligence*, McGeogh's *Psychology of Human Learning*, Murchison's *Handbook of General Experimental Psychology*, Pavlov's *Conditioned Reflexes*, and Young's *Emotion in Man and Animal*.

Psychological Service Center

H. MELTZER

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 45

March, 1954

Number 3

UNRELIABILITY OF GROUP TEST PROFILES¹

FLORENCE T. SNODGRASS

University of New Brunswick

The literature on trait variability within the individual is slight compared with that on individual differences. However, with the current interest in profiles and patterns, there is likely to be more attention paid to the subject in the future. Formerly, the emphasis in measurement was on the level of performance, but the mounting attention to profiles has broadened the focus to include, also, the pattern of performance. This statement applies to the use made of both group and individual test results for purposes of guidance, diagnosis and prediction.

The wide range of individual variability reported by Hull² was corroborated by the later investigations of Foran, Lillie and O'Leary,³ and of Stout.⁴ And other studies, only incidentally concerned with extent of variability, have all concurred with the findings of these three investigations. The variability of subtest scores in respect to the individual's mean intelligence test score is called unevenness. This variability or unevenness is described numerically by some type of differential score and is represented graphically by a profile. Schlessner's study⁵ is the only one, to the

¹ A summary of a dissertation prepared under the direction of Professor J. W. Tilton and submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Yale University, June, 1949.

² Clark L. Hull, *Aptitude Testing*. Yonkers-on-Hudson N. Y.: World Book Company, 1928, p. 47.

³ T. G. Foran, Gerald A. Lillie and Charles E. O'Leary, *A Study of Trait Variability*. Educational Research Bulletins, Washington, D.C.: The Catholic University Press, 1928, p. 23.

⁴ H. G. Stout, "Variations of normal children," *Journal of Experimental Education*, 6: 84-100, September, 1937.

⁵ George E. Schlessner, "Development of special abilities by the juvenile high age," *Journal of Educational Research*, 40: 39-51, September, 1947.

writer's knowledge, which has dealt with the reliability of this profile. In comparing the reliabilities of ability scores with the reliabilities of the differential or unevenness scores obtained from an achievement test battery, Schlessor found that even in the most reliably tested ability, slightly less than twenty per cent of the trait differences could be said to have significance.

Little work has been done to describe how stable the profile is from one test performance to the next. There may be much irregularity in the effort put forth by individual pupils in taking a group intelligence test. Some may work hard on one subtest and let up on the next; others may work consistently throughout the test battery. Such irregularity of effort would modify the shape of the profile essentially caused by trait differences. These possibilities raise two questions:

1) To what extent is the shape of the profile related to the order of subtests?

2) Is the shape of the profile related to the interval between subtests?

The relation between the extent of unevenness and educational achievement has been studied. DeVoss⁶ found only slight and insignificant differences between the unevenness of abilities of gifted and unselected children. Gray⁷ found individual variability in educational achievement somewhat greater in the dull group than in either the average or bright group, but agreed with DeVoss in finding no true difference between the average and the bright. Measuring achievement by college marks instead of a standardized test, Rausch⁸ found students with the greatest variability to have the poorest achievement in college. Pressey⁹ found no evidence of any real difference in the amount of unevenness between the dull, average and bright. Merrill,¹⁰ on the other hand, concluded that

⁶ L. M. Terman, Editor, *Genetic Studies of Genius*. Stanford University, Stanford University Press, 1925, chapter 12.

⁷ Susan W. Gray, "The relation of individual variability to intelligence," *The Journal of Educational Psychology*, 35: 201-210, April, 1944.

⁸ Oscar P. Rausch, "The effects of individual variability on achievement," *The Journal of Educational Psychology*, 39: 469-478, December, 1948.

⁹ S. L. Pressey, "Irregularity on a Binet examination as a measure of its reliability," *Psychological Clinic*, 12: 236-240, May, 1919.

¹⁰ Maud A. Merrill, *On the Relation of Intelligence to Achievement in the* of M. Retarded Children. Comparative Psychology Monographs, 10.

unevenness was greatest for the superior group, less for the dull and least for the normal. Brown¹¹ found about equal unevenness in three special abilities for the dull and bright. Woodrow,¹² holding mental age constant, found that unevenness was at a minimum for children of average IQ and that children of either low or high IQ have greater unevenness than those of average. He also found a tendency, though not one with a high degree of reliability, for the children of low IQ to be more uneven than those of high IQ. Tilton¹³ disagreed with these conclusions after making a critical analysis of Woodrow's data. From data of his own, Tilton made the same inferences that resulted from his re-examination of Woodrow's data; namely, an inverse relationship between unevenness and IQ. In a later study Tilton¹⁴ found that unevenness was also directly related to progress in school. Pupils accelerated for their mental age were more uneven than the at-grade group, and the retarded pupils were less uneven than the at-grade. These findings led Tilton to suggest that these relationships might be due to unreliability and to raise these questions: Is dullness associated with more than average unreliability in taking group intelligence tests? Is the greater unevenness of the accelerated group at school a function of unreliability? Another direct relationship which Woodrow¹⁵ reported was between increased unevenness and advancing mental age. Might this direct relationship, also, be a function of unreliability?

In light of the issues that have been raised in the preceding paragraphs, this study was undertaken to answer the following questions:

- 1) How reliable are profiles?
- 2) Do slight variations from standard procedure in the administration of tests alter the profiles?
- 3) Are individual differences in unevenness due to individual

¹¹ A. W. Brown, *The Unevenness of the Abilities of Dull and of Bright Children*. New York: Teachers College Contributions to Education, No. 222, 1926.

¹² Herbert Woodrow, "Mental unevenness and brightness," *Journal of Educational Psychology*, 19: 289-302, May, 1928.

¹³ J. W. Tilton, "The relation between IQ and trait difference as measured by group intelligence tests," *Journal of Educational Psychology*, 38: 343-352, October, 1947.

¹⁴ J. W. Tilton, unpublished manuscript, 1948.

¹⁵ Herbert Woodrow, op. cit., p. 294.



differences in unreliability? If so (1) Is the greater unevenness of the dull due to their greater unreliability? (2) Is the increased unevenness in advancing mental age due to unreliability? (3) And is the greater unevenness of the accelerated at school due to unreliability?

PROCEDURE

The experiment was conducted in the junior high schools of New Haven, Connecticut. The test used was the Terman-McNemar Test of Mental Ability, Form C, which is a battery of seven subtests. It was administered in all the schools in the usual manner. A week later it was repeated in the same way to approximately one-third of the population, to be known in this study as the standardized group. After the same interval it was administered, by reversing the order of the subtests, to another third to be known as the separated group, by spacing each subtest on consecutive school days. The interval of one week was thought to be short enough to obviate marked gains due to learning, and long enough to minimize practice effects. The experiment was so designed in order to determine if the unreliability affecting profiles is changed by these two variants in standardized administration.

Since Woodrow¹⁶ had found a marked increase in unevenness with advancing mental age, it was necessary in this study to hold mental age constant in order to keep this relationship under control. The range of mental age levels selected for the sample was dependent on those levels for which the test was adequate. When the distributions of the various mental age groups were tested for skewness, the 12:0-12:5 was the lowest age group and the 14:6-14:11, the highest found to be without skewness. Accordingly, the sample selected included the six half-year mental age groups from 12:0-14:11, inclusive, which numbered 1138 cases. At each of these levels each subtest distributed the members of the group satisfactorily and without skewness.

Inasmuch as it was necessary to express each subtest score in comparable units for each of the six mental age groups, all scores were converted into sigma scores. For this purpose, the total sample was divided both on the basis of sex and half-year mental age levels. The terms unevenness score, differential score and trait

difference score are used interchangeably to mean the difference between the individual's average sigma score and his sigma score on a particular subtest. Hence the average unevenness score for each individual was the mean deviation of that individual's seven scores.

The unreliability referred to in this study is the inconsistency of test performance from one administration of the test to the next, using the same form. To measure the unreliability of each pupil's performance a score known and defined in this study as a discrepancy or 'D' score was adopted. D scores were derived in two different ways: one set from raw scores, one from sigma scores. A raw D score was defined as the sum of the differences, after correction was made for practice effects, between the raw scores on the first and second administration of the test. The D scores were derived from sigma scores by the same method except that there was no necessity to correct for practice effects. It may be noted that the direction or sign of the difference is of no importance in measuring the individual's unreliability. The larger the individual's D score, the greater is his unreliability in taking the test.

THE RELIABILITY OF THE PROFILE

Since the profile may be described with unevenness scores, its reliability can be measured by them. It is of course recognized that part of the profile is due to the unreliability of measurement. This fact is often ignored and the profile is misused because too much faith is put in its precision. The question arises: how high a peak or how low a dip must there be in a profile before we can be certain that it indicates an ability or disability? The reliabilities of the unevenness scores, which measure trait differences in the individual, are compared with the reliabilities of corresponding scores from which the profiles were made, as shown in Table I.

The first column in this table gives the reliability of the abilities as measured by the seven subtests from the two administrations of the test battery. An ability will be defined as the standard score a pupil makes on test 1, 2, 3, etc. The average of each pupil's standard scores will be called the pupil's general level of ability. Special ability or disability will be defined as his standard score on a given test minus his general level or average. Thus the peaks or dips of the profile above and below the general level will be graphic representations of special abilities and disabilities,



TABLE I.—COMPARISON OF THE RELIABILITIES OF TRAIT AND TRAIT DIFFERENCE SCORES FOR THE STANDARDIZED GROUP

Subtest	Reliabilities of Abilities	Reliabilities of Trait Differences	Standard Error of Measurement of Trait Difference Scores
Q 1	.64	.62	.683
2	.32	.29	.860
3	.72	.50	.566
4	.54	.43	.782
5	.78	.61	.657
6	.47	.34	.852
7	.66	.51	.783
	av. z = .707126 av. r = .605	av. z = .522644 av. r = .480	

spectively. And the term 'trait difference' heading the second column will be used to mean either a special ability or disability. (The reader is reminded of the interchangeable usage in this article of the terms trait difference score, differential score and unevenness 'U' score to measure these peaks and dips above and below the general level.) Here also the reliabilities from the two test administrations were calculated.

Schlesser¹⁷ reported a similar comparison of reliabilities of abilities and trait differences. Using the severe test of three standard errors of measurement from the mean before he admitted a reliable differential score, he found that, even in the most reliably tested area, only about twenty per cent of trait differences were significant. Applying this same criterion for significance to the present data, only about five per cent of the trait differences in the most reliably tested area (subtest 5) are significant. More general practice is to regard a critical ratio of 1.96 as being sufficiently high to indicate significance. An examination of the present data with this adjusted criterion, indicated that twenty-five per cent of the peaks and dips are significant in the most reliably tested area (subtest 5). In the least reliably tested area (subtest 2), ten per cent are significant.

From these findings it is concluded that most of the irregularities of a profile, as measured by the test used, are due to errors in meas-

urement. Profile differentiations are reliable only about twenty-five per cent of the time in the most reliably tested area.

THE RELATIONS OF UNRELIABILITY OF PROFILES TO TEST ADMINISTRATION

The extent to which the shape of the profile was related to the order of subtests was investigated by comparing the standardized group with the reversed group. The practice effects as indicated by raw scores were compared for these two groups and none of the differences were significant. Furthermore, the range and standard deviation on corresponding subtests for the two groups were practically the same. Additional evidence of the similarity in shape of the profiles of these two groups was found by comparing their average scores (measure of individual's unreliability or change in taking the test from one administration to the next) computed both on the basis of raw scores and sigma scores. In both cases the standard error of the difference far exceeded the obtained difference between the two groups. A comparison made of the reliabilities of the trait difference scores on each subtest showed no significant difference for the average of the two groups, nor did it show any consistent direction in the differences between the reliabilities of corresponding subtests. Surprisingly enough, subtest 4, the relative position of which was not changed in either group, showed the greatest amount of change.

A final comparison of these two groups was made of the reliabilities of their average U_1 and U_2 scores¹⁸ as shown in Table II. The unevenness scores from the two administrations of each subtest were correlated. Then these r 's were converted into z values for averaging and the average z 's converted back to r 's. It will be noted that the difference on the whole between the two groups was not significant, nor was there any consistency in the direction of the differences between coefficients of corresponding subtests, as Table II shows. The standardized and reversed groups were shown to be similar by all the criteria applied: average gain, measured both in terms of central tendency and spread; discrepancy or change from test to retest; reliability of both subtest scores and unevenness scores. The reliabilities of the subtests showed a criss-

¹⁸ The subscripts refer to the U scores obtained from the first and second administration.

TABLE II.—COMPARISON OF RELIABILITY COEFFICIENTS OF UNEVENNESS SCORES FOR THE STANDARDIZED AND REVERSED GROUPS

Subtest	Standardized Group	Reversed Group	SE of Difference
1	.62	.53	
2	.29	.44	
3	.50	.54	
4	.43	.49	
5	.61	.56	
6	.34	.36	
7	.51	.50	
	av. z = .52264	av. z = .53737	.02598
	av. r = .480	av. r = .490	C.R. = .566

crossing in the direction of the differences, which was hard to account for, from the hypothesis set up. Also, the test showing the greatest amount of change was the one whose relative position was the same for both groups; the reliability of scores on the first and last subtests, on the average, does not seem to have been affected by the reversal. For these reasons it would seem that it was not the reversing which caused the changes, but other unaccounted for factors. Hence it may be concluded that the order of the subtests is not related to the shape of the profile. Whatever unreliability there is in the profile is not due to the order of subtests.

Does a battery of subtests all taken at the same time, i.e., at one sitting, yield the same profile that would be obtained if the several tests were taken at different times? A comparison of the results for the standardized group with those for the separated groups provides an answer. When the practice effects of these two groups were compared none of the differences were found to be reliable. In addition, the two groups were practically the same in regard to range and standard deviation. Further evidence of the similarity in the shape of the profiles of these two groups was obtained when a comparison of their average D (unreliability) scores was made. The standard errors of the difference between these scores, computed on the basis of both raw score and sigma score data, far exceeded the obtained difference.

comparison of the reliabilities of subtest scores (trait meas-

ures) showed that while the average coefficient of the separated group was significantly lower than that for the standardized group, the direction of the differences was inconsistent. This, coupled with the fact that only two of the subtests were significantly lower in the separated group, led the writer to compare the separated with a larger group. Accordingly, the reversed group was combined with the standardized group since the two had been found not to be reliably different. Compared with this enlarged group the difference in average reliability for the separated group was not reliably lower, and again there was no consistent direction in the differences between the subtest coefficients for the two groups. Only one subtest (No. 5) was significantly lower in the separated group than in the combined group. Finally, the reliabilities of the average U_1 and U_2 scores of the separated and standardized groups showed no reliable differences; nor was there any consistency here in the direction of the differences between the reliabilities for the corresponding subtests, as examination of Table III will show.

To restate, the standardized and separated groups were similar with respect to the measurements described above. Hence it is concluded that the separation of the subtests (by the interval used) is not related to the shape of the profile. The unreliability of the profile is not the result of the interval between subtests. This finding may have practical implications for the user of tests: The administration of many tests requires a longer time than one class period. The tester need not feel hampered by the demands of such

TABLE III.—COMPARISON OF RELIABILITY COEFFICIENTS OF UNEVENNESS SCORES FOR THE STANDARDIZED AND SEPARATED GROUPS

Subtest	Standardized Group	Separated Group	SE of Difference
1	.62	.50	
2	.29	.38	
3	.50	.51	
4	.43	.33	
5	.61	.57	
6	.34	.36	
7	.51	.49	
	av. z = .522644	av. z = .487912	.03051
	av. r = .480	av. r = .450	CR = 1.12

time restrictions but can separate the battery to fit the time schedule.

RELATION BETWEEN UNRELIABILITY AND IQ

A third purpose of this study was to investigate the cause of individual differences found in unevenness. These data were first examined for the presence of the inverse relationship between IQ and unevenness which was reported by Tilton.¹⁹ The findings were adequate to justify a search for any possible relation between unreliability and IQ.

The reader is reminded of an earlier statement regarding the use in this study of a discrepancy score (D score) to measure the unreliability of each pupil's performance. For the total sample the reliability of the discrepancy scores was .08 with a SE of .028. (When this r is stepped up by the Spearman-Brown formula it becomes .15.) While an r of this size indicates a very low reliability without practical significance, it more than passes the one per cent level of statistical significance.²⁰ There is therefore a slight tendency for some individuals to be more unreliable than others. Are these more unreliable pupils the dull pupils?

The D scores for the dull, average and bright groups were as shown in Table IV.

In determining the probability that the dull and average will always exceed the bright in unreliability, only the differences in one direction are involved: that is, (1) excess of unreliability of the dull over the bright, and (2) excess of unreliability of the average over the bright. These differences are reported in Table V. A difference between the dull and bright of this magnitude (2.47) and in this direction passes the one per cent test. In the same way the difference between the average and the bright satisfies the two per cent test, and the difference between the dull and average combined, and the bright more than satisfies the one per cent test. And the findings are very similar when the discrepancy scores were computed from raw score data. Evidently the bright are more reliable than the dull and average. It has been shown therefore that the dull and average are not only more uneven, but are also more unreliable; and that the bright are not only less uneven, but also

¹⁹ J. W. Tilton, loc. cit.

²⁰ E. F. Lindquist, *Statistical Analysis in Educational Research*, Boston: Houghton Mifflin Co., 1940, p. 212.

TABLE IV.—DISCREPANCY SCORES FOR DIFFERENT IQ GROUPS (SIGMA SCORE DATA)

	IQ below 95	IQ 95-104	IQ above 104
M	63.54 (372)	63.19 (421)	59.63 (345)
SD	21.56	20.91	20.80
6M	1.12	1.02	1.12

TABLE V.—COMPARISON OF THE DIFFERENCES IN DISCREPANCY SCORES FOR DIFFERENT IQ GROUPS (SIGMA SCORE DATA)

	Difference	Standard Error of Differences	CR
Dull-average	.35	1.51	.23
Dull-bright	3.91	1.58	2.47
Average-bright	3.56	1.51	2.36
(Dull + average)-bright	3.72	1.35	2.76

less reliable. Hence, the negative relationship between unevenness and IQ which has been reported could be a function of individual differences in test-retest unreliability.

THE RELATION BETWEEN UNRELIABILITY AND MENTAL AGE

Stemming from the relation found between IQ and unreliability, a question arises as to the possible relation between unreliability and mental age. Using raw score data the average discrepancy scores were 13.48, 12.95 and 13.13 for the twelve, thirteen and fourteen mental age level, respectively. Although these scores tend to decrease with advancing mental age, they are not consistent in this direction; and the differences are not significant (critical ratios of 1.71 and 1.06, respectively). Hence these data provide no basis for attributing to unreliability the relation which Woodrow²¹ found between increased unevenness and advancing mental age.

Similar comparisons were made using the D scores derived from sigma scores. Here, however, IQ has been held constant lest the inverse relationship between IQ and unreliability, which has been found in the data, should complicate the findings. For the average IQ group these discrepancy scores are 65.01, 62.32 and 62.67 for the twelve, thirteen, and fourteen mental age levels, respectively.

²¹ Herbert Woodrow, *op. cit.*, p. 294.

These D scores also decrease with advancing mental age and in a pattern similar to that of the other D scores. The slight deviation from the general trend of the fourteen mental age level is not significant. In this case, also, the differences are not significant; critical ratios are 1.01 and .91, respectively. Accordingly, there is no basis in these data for thinking that the relationship which Woodrow reported was due to individual differences. In fact, the differences here are in a direction opposite to that which would be necessary to explain Woodrow's findings on the basis of unreliability.

RELATION BETWEEN UNRELIABILITY AND ACCELERATION IN SCHOOL

Reference was made in the Introduction to a direct relationship found by Tilton²² between unevenness and acceleration in school: the accelerated for their mental age being more uneven than the at-grade group, and the retarded less uneven than the at-grade. Could it be that the greater unevenness of the accelerated is due to a greater unreliability on their part? Since unevenness has been found to be a function of IQ, the relation between unevenness and acceleration in school should be studied by holding IQ and mental age constant. For this reason IQ was held constant between 95 and 104. This average IQ group (95-104) of each of the three mental age levels of the sample was divided into the retarded, at-grade and accelerated on the basis of their grade placement in school. On the basis of normal progress through school, those in the twelve mental age group were considered at-grade if they were in the seventh grade and accelerated if in the eighth or ninth; those in the thirteen mental age group were considered retarded if they were in the seventh grade, at-grade if in the eighth and accelerated if in the ninth; and in the fourteen mental age group, retarded if they were in the seventh or eighth grades and at-grade if in the ninth. It is obvious from the foregoing that only the thirteen mental age group here contained all three classifications. Because of the incompleteness of the twelve and fourteen mental age groups for this classification, with the resulting small number of cases in the retarded and accelerated groups; and because no consistent pattern of unevenness was revealed for mental age levels; all three were thrown together for analysis. The mean unevenness score for the average group (95 to 104 IQ) was 80.38 for the retarded group (N = 60), 83.11 for the at-grade group (N = 287)

²² J. V. Tilton, unpublished manuscript, 1948.

and 83.92 for the accelerated group ($N = 74$). While the differences were insignificant, they were in the same direction as reported by Tilton.²³ Here, as in the relation between unevenness and IQ, these data could not be expected to show as pronounced differences as Tilton's since they include only a three-grade range, while his data covered a five-grade one. In view of the foregoing, the data were considered suitable for an investigation of the relationship between unreliability and acceleration in school. It will be remembered that the other step preliminary to this investigation—namely, that the unreliability scores are statistically reliable—has been demonstrated in a previous section.

The same average IQ group (421 cases) was examined for unreliability with these results: retarded 13.58; at grade 13.37; accelerated 13.04.

The differences between the three groups were so small as to be of no statistical significance. (Critical ratios of .71, .87, and .60) It should be noted, too, that the trend of the differences in the reliability of these school groups is just opposite to that which one must find, in order to explain the direct relation which Tilton reported between unevenness and acceleration in school according to mental age.

Another comparison of the unreliability of the three school groups was made on the basis of the D scores computed from sigma scores, with these results: retarded 61.00; at grade 63.64; accelerated 61.93. Here, also, the differences were not significant (the standard errors being in each case larger than the obtained differences). In these data there is, then, no evidence that the relationship between unevenness and acceleration may be caused by unreliability. In other words, the greater unevenness in the profiles of the accelerated is not the result of their greater unreliability.

SUMMARY

The data seem to justify the following conclusions, within the limits of the test used and subject to the conditions of the investigation:

- 1) There is dependable evidence that much of the unevenness of the profiles is due to test unreliability. In the most reliably tested areas, only about twenty to twenty-five per cent of the time are the trait differential scores or unevenness scores reliable; and in the

²³ J. W. Tilton, loc. cit.

least reliably tested areas only about five per cent of the time are they reliable. These findings have implications for the guidance worker who should be cautious about considering all high and low differential scores as evidence of a special ability or disability.

2) The relation of the unreliability of profiles to test administration was negligible. Profiles were not changed significantly by the two variations of administration investigated. These findings may have implications for the makers and users of tests: a test battery may be broken up without fear of its effect on profiles.

3) Unreliability itself was found to have a very low but statistically significant reliability.

4) The unreliability of profiles was found to be inversely related to IQ, in that the bright were found to be more reliable than the dull or average in taking group intelligence tests. From this finding it seems reasonable to conclude that the greater average unevenness of the dull is caused in part at least by their greater unreliability.

5) The data provide no basis for attributing the direct relationship between average unevenness of profile and mental age, to unreliability.

6) And further, the data provide no basis for attributing the direct relationship between average unevenness of profile and acceleration in school for mental age to unreliability.

REFERENCES

1) Lee J. Cronbach, "Test reliability: its meaning and determination," *Psychometrika*, 12: 1-16, March, 1947.

2) Louis Guttman, "A basis for analyzing test retest reliability," *Psychometrika*, 10: 255-282, December, 1945.

3) George E. Schlessler, "Development of special abilities at the junior high school age," *Journal of Educational Research*, 40: 39-51, September, 1946.

4) J. W. Tilton, "The relation between IQ and trait difference as measured by group intelligence tests," *Journal of Educational Psychology*, 38: 343-352, October, 1947.

5) —, unpublished manuscript, 1948.

6) —, "An experimental effort to change the achievement test profile," *Journal of Experimental Education*, 15: 318-323, June, 1947.

7) F. T. Tyler, "Analysis of the Terman-McNemar Tests of Mental Ability," *Education and Psychological Measurements* 5: 45-58, No. 1, 1945.

8) Herbert Woodrow, "Mental unevenness and brightness," *Journal of Educational Psychology*, 19: 289-302, May, 1928.

STUDENT-CENTERED VERSUS INSTRUCTOR-CENTERED INSTRUCTION¹

W. J. McKEACHIE

University of Michigan

Even psychologists have their stereotypes. And for most of us 'student-centered' and 'instructor-centered' are stereotypes. With 'student-centered' we associate the halo terms of democratic, permissive, insight, affective and student growth. 'Instructor-centered' brings to mind the terms authoritarian, Fascistic, knowledge for its own sake, and content-centered. In our psychological subculture the mere labels in our title stack the deck against anyone who attempts to defend the instructor-centered point of view.

Despite our preconceptions, there are important areas of agreement; e.g., everyone agrees that we're teaching students and that our job is to promote student learning. Also, almost everyone agrees that we want to improve our students' problem-solving skills. This paper aims to do three things:

- 1) To break up our stereotypes by examining some of the dimensions which may differentiate student-centered from instructor-centered instruction.
- 2) To survey a group of research studies on the problem.
- 3) To do some theorizing about the problem.

DIMENSIONS OF DIFFERENCE

First, what are the dimensions of difference?

GOALS

One of the most prominent is the dimension of goal setting. The instructor-centered teacher believes that he is ultimately responsible for determining goals. To quote the report of a study group which met at Cornell last year, "If the teacher merits the responsibility placed in his hands, he knows more than do the students about the subject, about the world in which we live, and of the

¹This paper was presented in an APA symposium, "Student-centered versus Instructor-centered Instruction," sponsored by Divisions 2 and 15 with Dr. Percival Symonds as chairman, Sept. 2, 1952.

ways in which a knowledge of psychology can enrich the world" (8). The student-centered instructor, on the other hand, believes that the group, including both students and instructor, should determine the group goals.

Despite his greater emphasis upon the student in goal-setting, the student-centered instructor usually has certain implicit goals which he hopes will be achieved by the students. Thus another area of difference between instructor-centered and student-centered teaching is in the type of goals for which they are aiming. This area of goals needs to be plotted in at least two dimensional space because there are three differing types of goals: intellectual, applied, and affective.

One approach emphasizes the traditional intellectual goals of a liberal education. It attempts to create an interest in 'Knowledge for its own sake'. The primary goal is to teach students to think. The instructor with this approach may be interested in attitudes, but they are the attitudes of the scientist toward his subject matter, not social attitudes. These are the goals which we usually associate with the instructor-centered method, but which may be independent of other dimensions of the methods.

The instructor who is interested in applied goals feels that psychology can offer much that is useful in the student's daily life, not only in adult life but in college as well. He is interested in teaching the facts of psychology only insofar as they can be applied by the student. Specific skills such as reading skills or study habits are apt to be part of the content of his course.

The instructor primarily interested in affective goals is likely to disavow any suggestion that his class is simply group therapy. In spite of this, he is even more dissatisfied than most teachers with achievement tests as a criterion of his teaching, and he is quite unhappy if his students don't have some adjustment problems that their psychology course can cure.

Student-centered instruction tends to emphasize the last of these three goal-emphases, but none of these goal-emphases are insolubly linked with other dimensions of teaching. Benjamin Bloom, for example, has been much interested in the discussion method as a technique for attaining intellectual goals.

METHODS OF TEACHING

Probably, most of us think of student-centered instruction as differing from instructor-centered teaching primarily in terms of

what goes on in the classroom. One of the dimensions here is degree of student participation. The student-centered instructor tends to encourage a high degree of student verbal participation. Most student-centered instructors, however, are not satisfied if student discussion is directed at the instructor. They are interested in developing a high degree of inter-student participation. The student-centered instructor feels that students who talk to him are maintaining their dependence upon him, and this is in conflict with his usual goal of independence.

Another dimension of classroom behavior is the degree of instructor acceptance of erroneous or irrelevant student contributions. Some student-centered instructors emphasize the importance of accepting all contributions without evaluation, or at least without negative evaluation.

A third dimension of classroom climate is degree of group cohesiveness. Typically the student-centered instructor attempts to create a group with a high degree of cohesiveness. He may attempt to measure this by counting the number of 'We's' as contrasted with the number of 'I's' verbalized in class discussion.

A further dimension upon which student-centered and instructor-centered classes differ is in the degree to which the student feels he can influence his own fate. Perhaps it is unnecessary to elaborate upon this point, but it may be pointed out that what have been called student-centered classes vary from classes in which the instructor lays out a course outline, makes assignments, and actively guides discussion, to classes in which almost all course planning is done by group decisions and the instructor begins class with, "What would you like to talk about today?"

The amount of class time devoted to personal experiences and problems of the students may be another dimension. The instructor who is concerned with the development of self-insight is more apt to permit or encourage discussion of personal problems.

Obviously one could list many other dimensions upon which instructor-centered and student-centered teaching may differ. Those listed are simply some which have been most frequently mentioned in the literature. When someone says he has used student-centered teaching he usually means that as compared with instructor-centered teaching, his class has had a higher degree of one or more of these qualities:

- 1) Student participation in goal setting.
- 2) Emphasis upon affective goals.

- 3) Student participation and student interaction.
- 4) Instructor acceptance of inaccurate statements.
- 5) Group cohesiveness.
- 6) Ability to determine its own fate.
- 7) Amount of time devoted to discussing personal experiences and problems.

EXPERIMENTAL RESULTS

One would expect that the controversy between our education's authoritarians and permissivists would long ago have been resolved by the cold logic of experimental studies. Unfortunately, this just hasn't happened. The published experimental studies are not in agreement and there are a host of unpublished studies which remain unpublished because the two methods used produced no significant differences in outcomes.

By way of some examples: One of the best known comparisons of student-centered and instructor-centered instruction is that of Faw (2). Faw's class of one hundred and two students met two hours a week to listen to lectures and two hours a week in discussion groups of thirty-four. One of the discussion groups was taught by a student-centered method, one by an instructor-centered method, and one group alternated between the two methods.

As compared with the instructor-centered class the student-centered class was characterized by more student participation, no instructor correction of inaccurate statements, lack of direction, and more discussion of ideas related to personal experiences.

Surprisingly enough, Faw's major measure of attainment of objectives was in the intellectual area. Scores on course examinations showed small but significant differences favoring the student-centered method. In the area of his major interests—emotional growth—Faw's method of evaluation was to ask students in the student-centered and alternating classes to write anonymous comments about the class. Generally these comments seemed to indicate that the students felt that they received greater social and emotional value from the student-centered discussion groups than they would have from an instructor-centered group. Despite the objective test results Faw's students felt that they would have made greater intellectual gains in an instructor-centered class.

Now compare Faw's experiment with that of Asch (1). Asch, like Faw, taught all of the groups involved in his experiment.

Three sections of about thirty to thirty-five students were taught by instructor-centered methods; one section of twenty-three students was taught by a student-centered method, quite similar to that of Faw. However, there were certain differences between Faw's and Asch's experiments. In Faw's experiment both student-centered and instructor-centered classes spent two hours a week listening to lectures. In Asch's experiment, only the instructor-centered classes were subjected to lectures. While Faw doesn't mention grading, one assumes that grades were determined by the instructor on the basis of the course-wide examination. In Asch's experiment students in the student-centered class were allowed to determine their own grades.

The interesting thing is that Asch's results do not completely agree with Faw's. On the final examination in the course students in the instructor-centered classes scored significantly higher than members of the student-centered class, not only on the objective portion of the test but also on an essay portion. Note, however, that the student-centered class was specifically told that the examination would in no way affect their grades in the course so that these differences may be simply due to a difference in motivation. As measured by the Bogardus Social Distance scale, attitude change in the two sections was not significantly different. However, as compared with the instructor-centered classes a greater percentage of members of the student-centered class improved in adjustment as measured by the MMPI.

Interestingly enough Asch's students, like Faw's, had a different perception of their achievement than that shown by the course examination. Faw's student-centered class did better on the course examination than the instructor-centered section, but thought they would have learned more if they had been in an instructor-centered class. Asch's students rated the student-centered class higher than instructor-centered in helping them to learn the subject matter of the course but they actually scored lower. There seems to be some irony in the fact that advocates of student-centered methods find that students' perceptions of group achievement is erroneous; yet many of us want students to take a larger share of the responsibility for evaluation, and are pleased that they report great gains in personal and social values in student-centered classes. If groups which report greater intellectual gains actually learn less, it might seem logical to conclude that groups which report greater gains in

the areas of personal emotional growth may really be gaining less than those groups which report lesser gains.

One of the most comprehensive experiments in this area is that of Landsman (5). He experimented with a student-centered type of teaching as contrasted with a more directed type of democratic discussion organized around a syllabus. His experimental design involved eight classes in a course sequence of Human Development, Adjustment, and Learning. Three instructors took part in the experiment and each instructor used both methods. Outcome measures included the Horrock-Troyer tests, a local case history analysis test, Group Rorschach, MMPI, autobiographies, and students' reactions. His results showed no significant differences between methods.

A REDEFINITION OF THE PROBLEM

What are we to conclude from these studies? While there is a dearth of follow-up data, with such slender results at the end of the courses our hope that either method produces significantly greater long-time benefits is probably unrealistic. Should everyone go his own way and teach any way he pleases? Personally, we are not willing to go quite so far, but certainly none of us should exclaim with horror, "His classes are instructor-centered!"

As psychologists, however, we believe in research. Why has research on student-centered versus instructor-centered teaching seemed to lead up a blind alley?

One reason suggested for contradictory results is that different people have meant different things by student-centered. But a far more important reason is that we've been lumping together more variables than we could handle with our experimental designs. The first part of this paper was concerned about defining dimensions of difference because we need to work with a more limited number of variables and we need to relate these variables to the main body of psychological theory.

Do we have any assurance that in the welter of student-centered versus instructor-centered research there are any variables that make a difference? Despite a somewhat pessimistic view of most research in this area, there are enough glimmerings of hope to justify further research.

For example, Smith and Johnson (7) have found that student-centered teaching produces higher scores than instructor-centered

teaching on tests of reasoning ability and creativity. Furthermore, all research on the problem seems to agree that as compared with instructor-centered teaching, student-centered teaching results in little decrement to the learning of facts (providing the classes have textbooks and tests are based on the texts).

In addition Asch and Faw are not alone in feeling that adjustment and social skills are improved by student-centered methods. The research of Gibb and Gibb (3) indicates that students from group-centered classes which possessed many of the characteristics ordinarily called 'student-centered' actually produced a growth in social skills in experimental situations outside the classroom. Kelley and Pepitone (4) found an increase of empathy in classes which had been taught by student-centered methods. Efforts to produce such gains in instructor-centered classes have been unsuccessful.

Obviously, we cannot test every possible interrelationship of the variables involved in student-centered classes, but many can be studied independently of the total method. For example, the variable of degree of student verbal participation in the classroom could be divorced from goals, cohesion, and other variables, and conceptualized as providing an opportunity of student responses to be rewarded or punished. We might gain some insight into student-centered teaching by varying opportunities for verbal participation or by varying the percentage of comments which the instructor rewards or corrects.

Another variable upon which we already have data is the degree to which the student feels able to influence his own fate. It seems significant that almost every one who tries student-centered teaching methods finds that the problem of grades presents the greatest obstacle to success. In addition two studies (3) (7) which produce evidence of change as a result of student-centered teaching both reduced the power of the instructor by giving the group responsibility for determining grades.

Our own research at the University of Michigan supports our notion that the student's feeling of freedom is an important variable. In an earlier paper (6), the writer reported that students made higher scores on examinations when they were given the opportunity to write comments about test items. More recent experimentation indicates that the effect of this opportunity is reduced if the student is directed to write comments. Apparently the important thing is that the student feel free to do so if he wishes to.

In addition, it was found in another study that students preferred a directive method of teaching which made clear what the student had to do in order to pass the course. Again, we would interpret this as meaning that the student felt better able to determine his fate.

Looked at in this way, the class in which the instructor insists on being non-directive may actually increase the student's feeling of helplessness since he doesn't know what to do in order to achieve the goal of a good grade in the course. In this situation the student may simply perceive the instructor as using his power to block the normal pathways to the goal.

From this theory one would predict that the effect of instructor permissiveness would depend upon whether or not the group possessed the skills necessary to achieve their goals. In a new group the effect of instructor permissiveness may depend upon the presence or absence of individuals in the group who have had previous experience in working in democratic groups. If the instructor retains control of rewards, permissiveness with respect to means to the goal (such as assignments, classroom activities, etc.) may simply increase the ambiguity of the situation for the student and reduce student learning.

Obviously such speculations need to be tested. To make our research productive we need to sharpen both our theories and our methodology.

REFERENCES

- 1) M. J. Asch, "Nondirective teaching in psychology." *Psychol. Monogr.*, 1951, 65, 4.
- 2) V. Faw. "A psychotherapeutic method of teaching psychology." *Amer. Psychologist*, 1949, 4, 104-109.
- 3) Lorraine Gibb and J. Gibb. *The Effect of the Use of "Participative Action" Groups in a Course in General Psychology*. Paper at 1952 APA meetings.
- 4) H. Kelley and A. Pepitone. "An evaluation of a college course in human relations." *J. educ. Psychol.*, 1952, 43, 193-209.
- 5) T. Landsman. *An Experimental Study of a Student-centered Learning Method*. Unpublished doctor's dissertation, Syracuse Univer., 1950.
- 6) W. J. McKeachie. "Anxiety in the college classroom." *J. educ. Res.*, 45, 1951, 153-160.
- 7) H. C. Smith and D. M. Johnson. *An Experimental Study of Attitudes and Achievement in the Democratic Classroom*. Paper read at 1952 APA meetings.
- 8) D. Wolffe, et al. *Improving Undergraduate Instruction in Psychology*. New York, Macmillan, 1952.

AN IMPROVED FORMULA FOR SCORING CERTAIN GUESS-WHO RATINGS AT THE ADOLESCENT LEVEL

EVAN R. KEISLAR

University of California at Los Angeles

The guess-who technique is a device whereby bi-polar trait ratings of a group of children may be obtained by using the members of this same group as the raters. It appears to be one of the few promising tools which yield information regarding the way each member of a large adolescent group is perceived by this group. But different assumptions and scoring methods have been used by investigators with the result that some ambiguity exists regarding the interpretation of scores yielded by this device. It was the aim of this study to develop a new scoring formula and to evaluate it in comparison with the two formulas previously applied when the guess-who technique is used for certain types of traits at the adolescent level.

SCORING METHODS OF GUESS-WHO RATINGS

In the usual form of this device, one or more pairs of statements are used for each trait, one statement of each pair representing behavior falling on what is arbitrarily designated as the positive side of the trait continuum, the other statement on the negative side. Raters are asked to indicate those members of the group who are judged to exhibit the behavior described by each statement. The CEI (2, pp. 221-36) used the following scoring method, referred to as Formula A, for such ratings:

Formula A.—Trait score = (No. of positive mentions) - (No. of negative mentions).

When the guess-who ratings are obtained from adolescents, where the peer group is of necessity large, there is a wide variation in how well different members are known by the raters. Consequently, Formula A may give excessively extreme scores for individuals who are widely known and scores which are probably too close to zero for those who are not. In most cases this factor of being widely known is not related to the trait in question.

To avoid this deficiency Havighurst and Taba (3, p. 21) used

what will be referred to as Formula B:

Formula B.—Trait score =

$$\frac{\text{No. of positive mentions}}{\text{Total No. of pos. and neg. mentions on the trait}} \times 100$$

But this formula appears to overcompensate, since it assigns the maximum score to each individual who receives only positive mentions regardless of their number. Furthermore, it requires discarding all cases which receive no mentions on the trait; many such persons may be well known but their behavior may be judged consistently by all raters to fall between the two statements.

Faced with such problems of scoring, Kuhlen and Lee (4) devised what amounted to two categories, one containing all scores with a greater number of positive mentions and the other consisting of all remaining scores. Bonney (1) translated the guess-who ratings to an arbitrary five-point scale. But such methods offer no adequate solution.

To attack this problem of scoring guess-who ratings obtained from large populations, let us assume that each pair of statements for a trait represents behavior falling beyond two points at approximately equal distances in opposite directions from the mean on the trait continuum. We may then assume that the number of mentions a person receives on a statement is equal to the product of two variables: (1) the amount of deviation shown by the individual on the trait in the direction indicated by the statement, and (2) a measure of how widely known the person is, that is, the number of raters who are acquainted with him. As an estimate of this second variable one may use the number of times an individual is mentioned by the raters on a wide variety of unrelated traits. By algebraically adding such deviations, after transposition, we obtain:

Formula C.—Trait score =

$$\frac{(\text{No. of pos. mentions}) - (\text{No. of neg. mentions})}{\text{Total No. of mentions on a variety of traits}} \times 100$$

This new formula, Formula C, is applicable only to traits rated on an instrument such as Tryon (5) developed in which a variety of trait ratings are obtained. It essentially yields trait scores which

would be expected if all individuals being rated were mentioned the same number of times on the entire set of traits. It would appear to be an improvement over Formula A, since it weights the number of mentions in inverse proportion to how widely known an individual is. It gives a score, as Formula B does not, to an individual who is not mentioned on the trait provided he is mentioned on others. An empirical comparison of Formula C with each of these other two formulas is therefore in order.

STATEMENT OF PROBLEM

It was the purpose of the present study to compare the proposed Formula C with Formula A and with Formula B, as previously defined, on two traits judged to be independent of this factor of being widely known, when the guess-who technique is used with secondary-school pupils. The bases of comparison were: (1) form of distribution yielded, (2) reliability of scores, and (3) validity of scores as measured by an independent criterion.

SUBJECTS

Guess-who ratings were obtained from one hundred and ninety-eight girls and one hundred and ninety-six boys out of the four hundred and thirty-six members of the B11 class in a senior high school located in a largely residential section of Los Angeles. These pupils were asked to rate the members of this class who were of their own sex. The group for whom ratings were scored consisted of the one hundred and seventy girls and one hundred and eighty-seven boys who had been enrolled a full school year.

THE GUESS-WHO DEVICE USED

Two forms of the guess-who rating blank were constructed, one for each sex. One pair of statements was used for each of twelve traits many of which were taken without major change from Tryon (5). The traits may be described by indicating, in the order in which they were listed on the rating blank, the opposing statements as follows: talkative—silent; old acting—young acting; friendly—unfriendly; likes school work—dislikes school work; considerate—inconsiderate; popular—unpopular (with opposite sex); not persistent—persistent; welcomed—ignored (by same sex); puts studies last—puts studies first; conceited—not conceited; cheerful—sad,

influential—not influential (girls form only); athletically competent—athletically incompetent (boys form only). Although, for the requirements of Formula C, traits could have been selected for the instrument which showed greater independence of each other, the variety appeared adequate.

TRAITS SELECTED FOR SPECIAL STUDY

Comparative data for the three formulas were obtained from the two traits 'likes schoolwork' and 'puts studies first' which are thus indicated by using only the statement assigned positive values. The wording of the statements for these two traits on the boys form, with appropriate changes for the girls in parentheses, was as follows:

Here is someone who likes schoolwork; he (she) usually enjoys what he (she) does in his (her) classes:

Here is someone who dislikes schoolwork; he (she) usually detests what he (she) does in his (her) classes:

Here is someone who will not do other things until he (she) finishes his (her) studying:

Here is someone who always spends his (her) time on other things even though he (she) has studying to do:

ADMINISTRATION AND SCORING PROCEDURE

In giving their ratings pupils were instructed not to sign their names and to indicate by means of pre-assigned code numbers those individuals they wished to nominate for each statement. The number of self-mentions, discovered by an invisible fluorescent ink device, was so small that they were not excluded in the scoring. The number of mentions received by the one hundred and seventy girls in the study on all twelve traits ranged from 5 to 375 with a mean of 68 and a standard deviation of 48. For the one hundred and eight-seven boys these scores ranged from 7 to 379 with a mean of 59 and a standard deviation of 45.

No Formula C scores were given to those pupils who received less than twenty-four mentions for the entire set of twelve traits. No Formula B scores were given to those pupils who received less than two mentions for the trait in question. Comparisons between Formula C and each of the other two scoring methods were made on the population of those individuals who received scores from both formulas being compared.

RESULTS

Four distributions, one for each of the two traits described earlier for boys and girls separately, were obtained for each formula. The means for Formula A and Formula C ranged from -1 to $+1$, with standard deviations ranging from 5 to 11. For Formula B the means ranged from 43 to 56 with all standard deviations approximately 40. Correlations between Formula A and Formula C scores ranged from .77 to .82; those between Formula B and Formula C scores ranged from .84 to .89. Since these scores were obtained from the same set of ratings, it is obvious that the formulas are yielding somewhat different results.

1) *Form of distribution.*—All four of the distributions for Formula A were highly leptokurtic. On the basis of the chi-square test, the hypothesis of normality could be rejected for each of these far beyond the .001 level. On the other hand all four of the distributions for Formula B were definitely bi-modal. Since about half of the scores were located at the extremes no test of normality was made. Although Havighurst and Taba used several pairs of statements for each trait, the bi-modal distributions usually found for this formula were taken as a sign of the reliability of the trait ratings (3, p. 217).

For Formula C, however, the distributions approximated normality. In three of these distributions the hypothesis of normality could not be rejected at the .05 level. For the fourth—liking of schoolwork by boys—the hypothesis could not be rejected at the

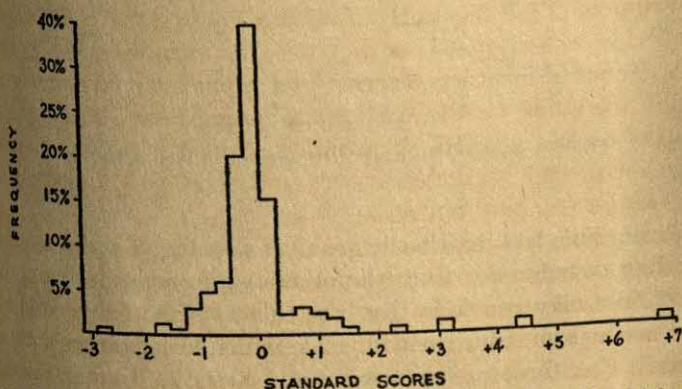


FIG. 1. Distribution of Formula A Scores on 'Likes Schoolwork' for 165 Boys.

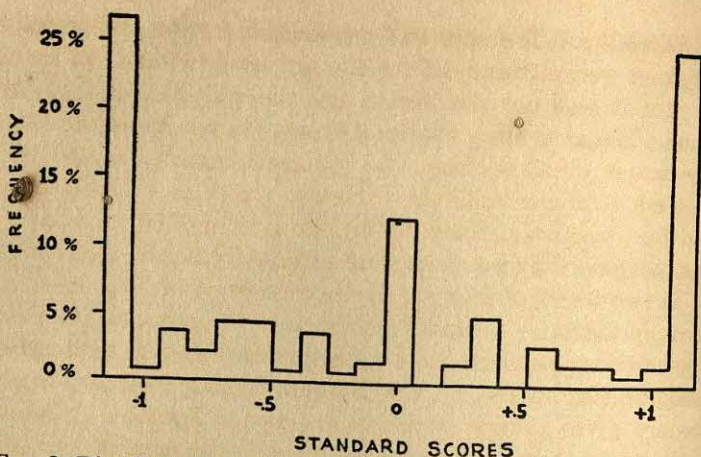


FIG. 2. Distribution of Formula B Scores on 'Likes Schoolwork' for 140 Boys.

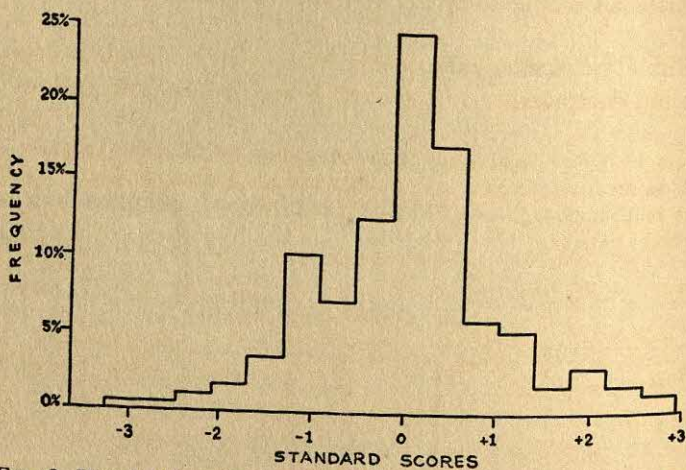


FIG. 3. Distribution of Formula C Scores on 'Likes Schoolwork' for 165 Boys.

.01 level. This last distribution is shown in Fig. 3 with the corresponding distributions from the other two formulas in Fig. 1 and Fig. 2. We may conclude that Formula C is the only one of the three methods of scoring which yielded distributions approximating normality on the two traits selected for study. Tests on most of the remaining ten traits in this instrument showed in every case that the new formula produced normal distributions.

2) *Reliability*.—The split-half reliabilities for the two traits for Formula A were .95 and .85 for the girls and .95 and .92 for the boys. But it was not possible to use the split-half method with Formulas B and C since the total scores are not merely the sums of the scores obtained from two groups of raters. Instead, since correlation coefficients between the two traits on Formula A corrected for attenuation were .98 for the girls and .99 for the boys, it was assumed the two traits were sufficiently similar to be looked upon as two forms of a rating device each of which was measuring approach-avoidance behavior with respect to schoolwork. Hence the correlations between the two traits were taken as estimates of the reliabilities offered by the different scoring methods. Such a procedure gives slightly spurious results for Formula C because the two sets of scores constitute indexes with the same denominator term. However, when the denominator (total number of mentions) was partialled out none of the correlations were reduced by more than .03.

In Table I the reliability coefficients for each of the two formulas being compared were obtained from identical populations as

TABLE I.—COMPARISONS OF RELIABILITIES OF SCORES YIELDED BY THREE FORMULAS. BASED ON CORRELATIONS BETWEEN TWO TRAITS, "LIKES SCHOOLWORK" AND "PUTS STUDIES FIRST" FOR BOYS AND FOR GIRLS SEPARATELY

Sex	Formula	N	r	Diff. bet. corresponding z's	σ_D^a	D/ σ_D
Boys	A	165	.93	.52	.11	4.67**
	C	165	.73			
	B	113	.74	.14	.14	1.04
	C	113	.80			
Girls	A	161	.86	.38	.11	3.41**
	C	161	.73			
	B	110	.68	.26	.14	1.90
	C	110	.79			

NOTE: All figures reported have been rounded off after computation to two decimal places.

^a Computed without considering a correlational term.

** Significant at .01 level

indicated by identical N's. The reliabilities given by Formula A are higher, both differences being significant at the .01 level, than those given by Formula C. This may be attributed in part to the highly leptokurtic nature of the distributions yielded by the first formula, thus allowing a few scores to contribute disproportionately to the covariance. Formula C gives estimated reliabilities which are

TABLE II.—COMPARISONS OF CORRELATIONS WITH SCHOOL GRADES AND SCORES YIELDED BY THREE FORMULAS. BASED ON TWO TRAITS FOR BOYS AND FOR GIRLS SEPARATELY

Trait	Sex	Scored by Formula	N	r with grades	Diff. bet. corresponding z's	D ^a /σ _D
Likes schoolwork	Boys	A	165	.61	.18	2.78**
		C	165	.71		
		B	140	.74	.05	.72
		C	140	.76		
	Girls	A	161	.61	.25	3.77**
		C	161	.74		
		B	133	.68	.22	3.17**
		C	133	.78		
Puts studies first	Boys	A	165	.57	.14	2.18*
		C	165	.66		
		B	126	.63	.10	1.75
		C	126	.69		
	Girls	A	161	.59	.16	2.66**
		C	161	.68		
		B	124	.70	.09	1.52
		C	124	.74		

Note: All figures reported have been rounded off after computation to two decimal places.

* All values of σ_D were either .06 or .07. These were computed from formula including the correlation term given by Quinn McNemar, *Psychological Statistics*, (John Wiley and Sons, New York, 1949), pp. 124-125. The eight values of r_{rr} ranged from .66 to .79.

* Significant at .05 level.

** Significant at .01 level.

higher than those yielded by Formula B, but these differences are not significant on the basis of the test used.

3) *Validity*.—The best validity criterion available for the two traits was grade point average during the pupils' previous year in high school. Although studying and liking of schoolwork are presumably not to be identified with good marks, there is evidence that the differences are not as large as might be expected. The partial correlations between each of the two traits, as scored by Formula C, and IQ, with grades held constant, are close to zero instead of being clearly negative. It would seem likely that the raters were highly influenced in their judgments by school marks pupils received or that teachers give grades partly on the basis of the interest shown by the student and completion of daily assignments, or perhaps both.

Grades were changed to a numerical scale where, for example, A = 9, D and F = 0. The mean grade for the girls on this scale was 4.7 with a standard deviation of 2.1. For boys the mean grade was 4.0 with a standard deviation of 2.4. The correlations between grades and the scores on both traits for the three formulas are given in Table II. As before, identical N's indicate identical populations.

The correlations for Formula C were significantly higher than those for Formula A in three cases. In the fourth the difference was significant at the .05 level. We may conclude that, if grades constitute an adequate criterion, Formula C yields higher validities for these trait ratings than does Formula A. Comparable data for Formulas B and C indicate that in each case Formula C yields higher validity coefficients, but only one of the differences is significant.

DISCUSSION AND CONCLUSIONS

1) Perhaps the most important finding in these comparisons is that the proposed formula was the only one of the three scoring methods which yielded normal distributions for these two traits. Although we have no evidence as to how the scores on a trait are distributed, normal distributions are to be preferred for practical reasons.

2) The general superiority of Formula C over Formula A with respect to validity would support the assumption, for traits and populations similar to those used in this study, that it is advisable

to weight in some manner the number of mentions on a guess-who statement in inverse proportion to a measure of how widely known the individual is among the rating group.

3) While the use of the total number of mentions on a variety of traits has given fairly satisfactory results as an estimate of extent of being known among the raters, a superior measure of this variable might be obtained by having raters check off on a list those members of the group whom they know. This would allow an investigator to use fewer traits with several pairs of statements for each trait.

SUMMARY

A new method for scoring guess-who ratings was proposed for those situations at the adolescent level (1) where the population being rated is so large that there exists considerable variation in how widely known by the group the members are, and (2) where ratings are being obtained on bi-polar traits which are judged to be relatively independent of this factor of being widely known. When guess-who ratings were obtained from a large high-school class on two traits dealing with approach-avoidance behavior with respect to schoolwork, the proposed formula was judged to be superior to each of the other two scoring methods which have been previously used. The new method of scoring produced higher validity coefficients with school grades as a criterion and was the only formula which yielded normal distributions. Some support was given to the assumptions made by the new formula.

REFERENCES

- 1) Merl E. Bonney, "Personality traits of socially successful and socially unsuccessful children," *J. Educ. Psychol.*, 1943, 34, 449-472.
- 2) Hugh Hartshorne, Mark A. May, and Frank K. Shuttlesworth, *Studies in the Organization of Character*, New York: The Macmillan Company, 1930, pp. 503.
- 3) Robert J. Havighurst and Hilda Taba, *Adolescent Character and Personality*. New York: John Wiley and Sons, Inc., 1949, pp. 315.
- 4) R. G. Kuhlen and B. G. Lee, "Personality characteristics and social acceptability in adolescence," *J. Educ. Psychol.*, 1943, 34, 321-40.
- 5) Caroline McCan Tryon, "Evaluation of adolescent personality by adolescents," *Monog. Soc. Res. Child Developm.*, 1939, Vol. 4, No. 4, Serial No. 23.

A COMPARISON OF MENTAL ABILITIES OF BRIGHT AND DULL CHILDREN OF COMPARABLE MENTAL AGES¹

OLIVER P. KOLSTOE

University of Illinois

The usefulness of the mental age as a unit of measurement depends upon several considerations. One is the extent to which children who achieve the same MA score are alike in their mental abilities. This investigation is concerned with one special case of this issue; namely, the question of whether or not there are systematic differences in performance on intellectual tasks between bright and dull children who achieve like MA scores; i.e., like general averages. Do MA's of 10, for example, achieved by bright seven-year-old children differ systematically from MA's of 10 achieved by dull thirteen-year-old children? Are the MA's earned by the two groups the same MA? Do we need to specify the age of a child in order to interpret his MA score?

Many investigators have tried to discover whether there are systematic qualitative differences between groups of children of varying chronological ages but of equal mental ages. In most instances the groups have been matched for comparison on the basis of mental age scores obtained from the administration of the Stanford-Binet scale. The investigators then identified the items of the Stanford-Binet which discriminated between groups of young children and older children equated on the basis of mental age. The item analysis data thus obtained were then presented as evidence of differences in the mental characteristics of the two groups.

This procedure is open to question for two reasons. In the first place, the criterion measures are not independent of those used for selecting and matching the groups. The requirement of equality of total mental age scores necessitates a balancing of differences between the two groups on the individual items of the test. Thus, superiority in one kind of task for one group necessitates superiority

¹ Based on the writer's doctor's dissertation, State University of Iowa, 1952, done under the direction of Professors A. N. Hieronymus and J. B. Stroud.

in another kind of task for the other group, since the total mental age scores for the two groups are equal. The fact that the selection and criterion measures are not independent makes it desirable that the differences observed in these studies be evaluated by a procedure in which the sources of differences are independent of the measures used for selection and matching.

In the second place, it is also questionable whether the matching procedure used in these former studies is defensible. When subjects are matched on the basis of obtained mental age scores, the estimated true mental age scores of the older, dull subjects are certain to exceed those of the younger, bright subjects, on the average. Thus, if groups are matched on obtained mental age scores, one would expect the low IQ group to obtain higher scores, on the average, than the high IQ group on intellectual tasks which are administered independently of the matching test.

The present study differs from those previously concerned with the problem in that (1) the low and high IQ groups were matched on the basis of estimated true mental age scores obtained by regressing their obtained scores; (2) tests used in the selection of the sample are not used as criterion tests; (3) the criterion tests were administered independently of those used for selection and matching and were more extensive than those used by other investigators.

RELATED RESEARCH

One of the earliest investigations of differences between old (dull) and young (bright) children having comparable mental ages was reported by Merrill in 1924. She concluded that the retarded group tended to excel on those types of items on which maturity, life age, and experience were favorable to success. In 1929 Wallin, finding the dull superior to the bright in the tasks of Counting Backward, and Vocabulary, pointed to age and experience as determining factors in this success. Laycock and Stanley in 1942 found the retarded group more successful in reproductive tasks, while the bright were superior in memory and reasoning. The same year, Rautman reported the bright superior to the dull in tasks involving manipulation. Rameseshan in 1949 found the bright superior on the Primary Mental Abilities of Verbal Meaning and Reasoning, whereas the dull were superior on Space. In examining the results of scores made on the California Tests of Mental Maturity Unsicker in 1950 found the bright superior on delayed recall

and reasoning, and the dull superior in spacial relations and numerical reasoning.

In the most extensive of the studies found dealing with the problem, Thompson and Magaret (1947 and 1950) found normal superior to dull children of comparable MA's in rote memory. In the second article they reported some evidence to indicate a special manual ability associated with mental deficiency.

PROCEDURE

Sample.—Two groups of children of approximately equal mental age but markedly different in chronological age and IQ were used in the study. A 'groups by levels' design was employed in the analysis of results. The groups consisted of twenty-nine third- and fourth-grade children with estimated true Stanford-Binet IQ's of 116 or above, and twenty-nine eighth- and ninth-grade pupils with estimated true IQ's of 84 or below. Four six-months intervals of estimated true mental ages, ranging from ten years, seven months to twelve years, six months, constituted the levels.²

The two groups were selected from classes in one mid-western city. Scores on the *California Tests of Mental Maturity*, which had been given previously to the children in this school system, were used to select pupils who appeared likely to fall into the two groups of bright and dull. The Stanford-Binet scale was then administered to the pupils so selected. Final selection of the subjects in the two groups was made upon the basis of the Stanford-Binet scores.

Estimated true scores used in constituting the two groups and MA levels within the groups were obtained by the use of standard regression equations. The reliability coefficients (Stanford-Binet Form L) used in the equations were: .982 for IQ's below 70; .945 for IQ's 70 to 83; .912 for IQ's 116 to 129; and .898 for IQ's 130 and above. Subjects in each group were assigned to levels on the basis of the estimated true mental age scores which resulted from the regression of the obtained scores. No further use was made of the Stanford-Binet results.

² A companion study by Bliesmer involving essentially the same samples as those used by the writer, was conducted to determine the nature of systematic differences in a number of reading abilities between bright and dull children of approximately equal mental age. Emory P. Bliesmer, *A Comparison of Bright and Dull Children of Comparable Mental Ages with Respect to Various Reading Abilities*, Doctor's Dissertation, State University of Iowa, 1952, pp. 153.

While the samples were not selected so as to control the number of boys and girls in each group, the sexes are represented in approximately equal numbers in each group.

Following is a diagram of the design as employed in the investigation:

MA Level	MA (Est'd. True)	Criterion Scores	
		Bright Group	Dull Group
I	12-1 to 12-6	n-7	n-7
II	11-7 to 12-0	n-10	n-10
III	11-1 to 11-6	n-7	n-7
IV	10-7 to 11-0	n-5	n-5

Measuring Instruments.—Used as criterion measures were the subtests of the *Wechsler Intelligence Scale for Children*, the *Benton Test of Visual Retention* (1945), an original *Speed of Symbol Copying* test, and the *Chicago Tests of Primary Mental Abilities* for ages eleven to seventeen. All of the criterion tests except the latter were administered in individual testing sessions. The Symbol Copying test was an adaptation of the WISC coding test, in which the subject was required only to copy the symbols presented in the top boxes into the bottom boxes. The tests were administered by four qualified examiners who worked independently. All of the tests were scored and rescored under the supervision of the writer. The results are reported in terms of simple raw scores.

Analysis of Results. For each criterion variable, the hypothesis tested is that there is no difference between the means of the populations of which the bright and dull groups are samples. The test of this hypothesis is the ratio of the mean square for groups to the mean square for within-cells.³

³ This ratio is a valid test of the hypothesis of equal population means only if the following conditions are satisfied: (1) The subgroup samples at each level are drawn independently at random from corresponding subpopulations. (2) The distribution of criterion measures in each subpopulation is normal and all have the same variance. (3) The numbers of measures in corresponding subgroups are in the same proportion for all groups and this proportion is the same as that among the numbers in the corresponding levels in the population.

There is no reason to question seriously the assumptions of randomness, normality, and homogeneity. The third condition listed above is definitely

A five per cent coefficient of risk was selected in advance of the analyses. Mean square ratios were considered significant if they equalled or exceeded the five per cent value in the F table. Throughout the investigation differences are reported as either significant or not significant at the five per cent level of confidence.

RESULTS

A summary of the differences between means of the groups for the various tests is given in Table I. In most cases the differences are small and not statistically significant, although some are significant.

The dull group was superior to the bright on the WISC subtests of Comprehension and Coding, on the Symbol Copying test scored for speed, and on the PMA subtest, Number. The bright group was superior to the dull on the WISC subtest Digits, and on the Symbol Copying test scored for accuracy.

It is suggested that the superiority of the dull group on the PMA subtest Number was due to the fact that this group, being made up of eighth- and ninth-grade pupils, had received instruction in certain arithmetic operations, specifically multiplication, to which at least part of the bright children, who were still in the third and fourth grades, had not been introduced. Thus, the superior performance of the dull group appears to reflect no qualitative intellectual difference, but rather, a curricular one. The results support a hypothesis of no difference between the two groups in the primary mental abilities.

Numerous investigators have pointed to a superiority of the dull

not satisfied in the present study since there is no possible distribution of numbers of subjects from level to level which can simultaneously be representative proportionally of the mental age levels in the populations of bright and dull subjects from which the samples were drawn. This lack of representativeness is of consequence only in the presence of a groups by levels interaction. There was little or no *apriori* basis for suspecting interaction (different group differences from level to level) in the present study, especially since each of the groups is fairly homogeneous with respect to age, mental age, and IQ. The significance of the groups by levels interaction was tested in each of the analyses by computing the ratio of the mean square for interaction to the mean square for within cells. In the event of a significant interaction the generalizing of the test of main effects (total-group differences) would be restricted to hypothetical populations showing the same proportions of subjects in the mental age levels as are found in the samples. None of the interactions were significant at the five per cent level.

TABLE I.—SUMMARY OF DIFFERENCES BETWEEN THE RAW SCORE MEANS OF THE BRIGHT AND DULL GROUPS FOR ALL CRITERION TESTS

	M _B	M _D	M _B - M _D	F
WISC				
Information	14.17	14.07	+	.04
Comprehension	12.24	13.90	-	1.66
Arithmetic	9.31	10.21	-	.90
Similarities	11.55	12.03	-	.48
Vocabulary	34.90	34.24	+	.66
Digits	9.93	9.00	+	.93
Pict. Comp.	11.97	11.28	+	.69
Pict. Arrange.	33.35	31.48	+	1.87
Blk. Design	22.48	23.07	-	.59
Obj. Assemb.	21.52	22.48	-	.96
Coding	36.10	48.55	-	12.45
Benton	12.28	11.83	+	.45
Symbol Speed†	91.45	141.07	-	50.62
Symbol Accur.	91.65	87.28	+	5.37
PMA				
Number	15.39	44.12	-	28.73
Verbal	30.23	30.46	-	.23
Space	15.42	15.69	-	.27
Word Fluency	33.23	33.35	-	.12
Reasoning	14.23	16.65	-	2.42
Memory	7.46	6.88	+	.58

* Significant at the 5% level of confidence.

† Symbol Speed scores are complements of time scores.

in tasks involving manual manipulation skills. The results of this investigation indicate a significant superiority of the dull group in performance on the WISC subtest Coding, and on the speed score of the Symbol Copying test. However, while it is true that the dull group did better than the bright on the speed part of the Symbol Copying test, when the test was scored for accuracy the bright group was superior. It is reasonable to believe that careful attention to neatness and accuracy, generally emphasized in the lower elementary grades, may have slowed down the speed of the bright group to a significant extent, thus leaving the relative superiority of the two groups on this subtest something of a question.

Moreover, the performance of the two groups on the WISC

subtests Completion, Picture Arrangement, Block Design and Object Assembly was not significantly different. These tests go to make up the performance part of the WISC. It would be expected that if the dull group were really superior in manual manipulation skill, they could be expected to be somewhat consistently superior to the bright group on tests which ostensibly require such skill. There are manual manipulative skills, performance on which correlate with criteria of intelligence, and those performances on which they do not. It is suggested here that if the dull older children are superior to the bright younger children on certain manipulative tasks, that these tasks may not be good intelligence tests in the first place.

The performance of the dull group on the WISC subtest Comprehension was significantly superior to that of the bright. An examination of the kinds of questions which are included in this subtest suggests that successful answers depend primarily upon experience, (for example; Why is it better to build a house of brick than of wood?).

The finding that the bright group was superior to the dull in their performance on the WISC subtest Digits supports the writings of other investigators who have reported a rote memory superiority for the bright groups. However, there is reason to doubt that they have any general superiority in memory ability since the small differences in favor of the bright group on the PMA subtest Memory and on the Benton test were not significant.

The results of the present study support to a very considerable extent the generality of the mental age concept. No evidence was found to support the claims of some writers that the bright are superior in such mental tasks as vocabulary, reasoning, and a general memory ability. Nor does the evidence from the study indicate superiority of the dull older children in performance tasks or manual manipulation.

Certainly this investigation has failed to reveal anything very wrong with the use of the Stanford-Binet MA as a measure of level of mental development.

REFERENCES

Periodicals—

- 1) Arthur L. Benton, "A visual retention test for clinical use," *Archives of Neurology and Psychiatry*, Vol. 54 (September, 1945), pp. 212-16.
- 2) Samuel R. Laycock and Clark Stanley, "The comparative perform-

ance of a group of old-dull and young-bright children on some items of the Revised Stanford-Binet Scale of Intelligence, Form L.," *The Journal of Educational Psychology*, Vol. 33 (January, 1942), pp. 1-12.

3) Ann Magaret and Thompson, Clare Wright, "Differential test responses of normal, superior and mentally deficient subjects," *The Journal of Abnormal and Social Psychology*, Vol. 45 (January, 1950), No. 1, pp. 163-164.

4) Maud A. Merrill, "On the relation of intelligence to achievement in cases of mentally retarded children," *Comparative Psychology Monographs*, Vol. 2 (September, 1924).

5) A. L. Rautman, "Relative difficulty of test items of the revised Stanford-Binet Scale; An analysis of records from a low intelligence group," *Journal of Experimental Education*, Vol. 10 (1942) pp. 183-94.

6) Clare Wright Thompson and Ann Magaret Thompson, "Differential test responses of normals and mental defectives," *The Journal of Abnormal and Social Psychology*, Vol. 42, (July, 1947), No. 3, pp. 285-93.

7) J. E. W. Wallin, "A statistical study of the individual tests in the Stanford-Binet Scale," *Mental Measurement Monograph*, No. 6 (1929), pp. 1-58.

Manuscripts—

1) Rukmini Selyan Rameseshan, *An Experimental Study of the Mental Age Concept*, Unpublished Master's Thesis, State University of Iowa, June, 1949.

2) Willard Daniel Unsicker, *A Psychological Study of Bright and Dull Children with Comparable Mental Ages*, Unpublished Doctor's Dissertation, State University of Iowa, 1950.

APPLICATIONAL TRANSFER AND INHIBITION¹

DONALD E. P. SMITH

University of Michigan

Progress is being made in formulating the relationships between variables belonging to different areas of psychological investigation (1, 2, 4, 6, 7). Miller and Dollard (2) have discussed the effect of repression upon ideation from the standpoint of learning theory. Others (1, 6) have reported data on the influence of affect on perceptual development.

The relation with which the present research is concerned is the effect of certain kinds of feeling-tone in the classroom learning situation upon 'applicable transfer' of learning. An achievement test was constructed for the purpose of measuring the student's ability to use information for solving problems in life-like situations. The position taken is that possession of information does not necessarily imply ability to use that information, but rather that such ability is partially dependent upon the amount and quality of emotional involvement in the learning situation and upon the personality structure of the learner. The extent to which products of past learning have been inhibited—one dimension of personality structure—is the subject of this paper.

The specific hypothesis to be tested is that limited ability to apply known principles in problem-solving is, in part, a function of the extent to which previously learned words and ideas have been inhibited or blocked. Applicable transfer is defined as a process by which relevant concepts are brought to conscious appraisal through the operation of necessary and sufficient generalization in the attempt to solve a problem. Inhibition as used herein refers to interference with the associative reaction in the normal operation of a 'cue-producing' response (2: p. 322), with a consequent reduction in efficient ideation. The cue-producing response which is blocked is viewed as capable of arousing feelings of fear

¹ A portion of the writer's doctoral dissertation entitled *Applicable Transfer: Its Nature, Measurement and Improvement* sponsored by M. D. Glock, Cornell University. The writer is indebted to Professors J. E. Hochberg, E. L. Walker, Ronald Lippitt and E. Lowell Kelly for critical reading of the manuscript.

or anxiety as a result of previous learning.² Symptoms of interference with the associative reaction, according to Rapaport (8: p. 9), are delay or absence of response, or repetition of the stimulus word in a word-association test.

The generalization process underlying applicational transfer is considered dependent for its sufficiency not only upon breadth of relevant meaningful experiences, but also upon relative absence of interference with necessary cue-producing responses. This view finds indirect support in Razran's report of studies in semantic generalization: "... generalization [of conditioned stimuli—single words] to the other word categories [synonyms, antonyms, etc.] was the greater ... the faster their reaction time in controlled association tests" (9: p. 355), whereas, as noted above, interference is characterized by delay or absence of reaction in a word association test.

The word association test was selected as a method appropriate for the measurement of inhibition as defined. An index of applicational transfer ability was provided by the differential between two scores on an achievement test in general psychology, reported elsewhere (11). The test consists of two parts, one purporting to measure information, the other, application, with each having eighty multiple-choice questions. The application part is composed of five reading selections concerning mental hygiene, learning, marital adjustment, all in story or anecdotal form, and two reports of experiments, followed by questions requiring identification of a principle, example given, of an example with principle given, of facts crucial to the solution of a problem, and others.

The problem is whether superior appliers and inferior appliers show a similar differentiation in respect to latency on a word association test, i.e., do superior appliers show brief latency and inferior ones delayed latency as might be predicted from the foregoing analysis.

SUBJECTS

Ss were fifteen introductory psychology students, each assigned to one of two groups on the basis of having a differential between his score on the application section of an achievement test in

² Dollard and Miller refer to this interference as 'repression' (2: p. 322) but the process, viewed empirically, also appears to fit O. H. Mowrer's formulation of the 'anticipatory response' (5: p. 131).

TABLE I.—CHARACTERISTICS OF SUPERIOR AND INFERIOR APPLIERS

Group	N	W-B Informa- tion	Word Fluency	Achieve- ment Test (total)	Applica- tion	Informa- tion	Differ- ence
Superior	9	19.1	57.7	75.3	33.1	42.2	9.1
Inferior	6	19.0	48.0	72.0	24.0	48.0	24.0
Difference		.1	9.7	3.3	9.1	5.8	14.9
P		—	.24	—	.05	.10	.01

psychology and his score on the information section sufficient to place him one SD above or below the mean of the group differential. The 'inferior appliers' ($N = 6$) scored, at the mean, 24 points lower on the application test than on the information test. (See Table I) 'Superior appliers' ($N = 9$) scored, at the mean, 9.1 points lower on the application than on the information test.

Despite the extreme differentials, there was no significant difference between the groups in total score on the achievement test in psychology, in scores on the *Weschler-Bellvue* information subtest, nor in scores on the *Thurstone Primary Mental Abilities* subtest of word-fluency. Those scores for the achievement test and for the information subtest are virtually identical.

APPARATUS AND PROCEDURE

The word association test has traditionally been used as an aid in determining inhibition (8: p. 35), conflict (3) or emotional disturbance (12: p. 363). Although qualitative analysis of responses is ordinarily used with individual cases, reaction time has yielded satisfactory results for groups (3, 10). One criterion of inhibition has been a latency greater than that of normal subjects. Gardner (3) reports a mean reaction time to 'neutral' words (e.g., tree, lamp, grass, salt) of 1.85 sec. for normal college students.

A group of fifty-seven words was adapted from a list of 'emotionally-toned' and 'neutral' words, so termed on the basis of 'practically universal agreement among competent judges', reported by Sharp (10: pp. 109-110). The words were chosen for their significance in several problem areas, viz., academic (study, exam), social (alone, friend), family (mother, father), and religious (sin, God). These were arranged with an attempt to alternate critical with neutral words, and were transcribed at 20 sec. intervals

on an Ampro Tape Recorder. The recording was utilized in order to increase the objectivity of the testing situation. Alternation with neutral words and use of intervals were introduced for the purpose of reducing the effect of possible perseveration due to emotional arousal.

Ss, who were tested 'blind' in a standard situation, were instructed to respond orally with the first word which came to mind after each stimulus word was heard, to respond to all words, and to attempt to keep the mind blank between words. Latency was recorded in fifths of a second by means of a stop-watch.

RESULTS

Mean seconds for each group are reported in Table II. The results clearly indicate differences favoring the superior applier group. All differences are significant beyond either the .01 or .02 level of confidence. Areas yielding the largest differences are religious and social; those yielding smallest differences, neutral and family.

DISCUSSION

It was predicted that, if the process of applicational transfer is influenced by inhibition of concepts involved in past learning, 'superior appliers' should show a lesser degree of inhibition, as determined by reaction time on a word association test, than should 'inferior appliers'. The predicted difference is clearly indicated.

Individual results are congruent with group means with the exception of one S in the inferior group who obtained mean latency below two seconds for all areas. Analysis of his test answer sheet

TABLE II.—DIFFERENTIAL LATENCY OF SUPERIOR AND INFERIOR APPLIERS ON A WORD ASSOCIATION TEST
(mean seconds)

Group	N	Related Area					
		Academic	Social	Family	Religious	Neutral	Total
Superior	9	1.66	1.84	1.25	1.44	1.56	1.57
Inferior	6	2.86	3.35	2.18	3.34	2.34	2.62
Difference		1.20	1.51	.93	1.90	.78	1.05
t		2.93	3.26	4.90	3.43	2.87	3.73
P		.02	.01	.01	.01	.02	.01

revealed that he failed to attempt twenty-eight of the eighty application items because, he subsequently stated, he was overly cautious on that part. Had he guessed on all omitted items, his application score, by chance alone, might have been such that he would not have been selected as a subject.

Rapaport (8: p. 17) suggests that a latency of four seconds be considered indicative of a disturbed response. Responses of inferior appliers include an average of 8.0 disturbed responses per subject whereas those of the superior group include an average of 1.5 such responses.

Alternative explanations may be offered. First, if reaction time is slower for the inferior group, it is possible that they did not complete the achievement test. If so, the differential between application and information scores might not be indicative of lesser application ability. The application section was administered first, however, and was completed by all well within the time limit.

Secondly, the inferior group may have been more cautious, thus omitting more items and lowering the application score. While this in itself might be of interest, an item count indicates that, with the exception of the one S discussed above, the poor appliers omitted less than half as many items per S as did the superior group. The possible significance of this result has not been explored.

Perhaps some unconscious bias on the part of the investigator, and consequent slowing of his own reaction time occurred even though the investigator was unaware of the group to which a subject belonged. Analysis of individual record sheets indicates that all but one of the inferior appliers had some words with latency comparable in brevity to those of the high appliers. Moreover, one member of the inferior group yielded a latency profile equivalent in brevity to those of the superior group except for one disturbed area. While not conclusive, these factors suggest the untenability of such an alternative.

SUMMARY

An attempt has been made to define 'applicational transfer' and to test the hypothesis that limited ability to apply principles in problem-solving is, in part, a function of the extent to which words have been inhibited or blocked. Fifteen Ss were assigned to two groups on the basis either of a large or a small differential between application scores and information scores on a previously validated

achievement test in psychology. The groups were shown to be approximately equated in certain other respects. With greater than normal latency on a controlled association test as the criterion of inhibition, superior appliers were found to have significantly shorter reaction time and fewer disturbed responses than inferior appliers. Alternative explanations are discussed.

A relationship between variables belonging to different areas of investigation, applicational transfer and emotional disturbance, has been suggested.

REFERENCES

- 1) J. S. Bruner and D. Krech, (Eds), *Perception and Personality: A Symposium*. Durham, N. C.: Duke Univ. Press, 1950.
- 2) J. Dollard and N. L. Miller, *Personality and Psychotherapy*. New York: McGraw Hill Book Co., Inc., 1950.
- 3) J. W. Gardner, "An experimental study of the Luria technique for detecting mental conflict." *J. Exp. Psychol.*, 1937, 20, 495-506.
- 4) J. L. Gewirtz, "Studies in word-fluency (I and II)," *J. gen. Psychol.* 1948, 72, 165-176.
- 5) O. H. Mowrer, *Learning Theory and Personality Dynamics*. New York: Ronald Press, 1950.
- 6) L. Postman, J. S. Bruner, and E. McGinnes, "Personal values as selective factors in perception." *J. abnorm. soc. psychol.*, 1948, 43, 142-154.
- 7) D. Rapaport, *Emotions and Memory*. Baltimore: Williams and Wilkins Co., 1942.
- 8) —, and R. Schafer, *Manual of Diagnostic Psychological Testing. Rev. Series, Josiah Macy Jr. Foundation*, III, 1946 (II).
- 9) G. Razran, "Stimulus generalization of conditioned responses." *Psychol. bull.*, 1949, 46, 337-365.
- 10) Delia L. Sharp, *Group and Individual Profiles in the Association-Motor Tests. Univ. of Iowa Studies in Child Welf.*, 1938, XV.
- 11) D. E. Smith and M. D. Glock, "Measuring knowledge and application: an experimental investigation." (To appear in *J. exp. Educ.*)
- 12) R. S. Woodworth, *Experimental Psychology*. New York: Henry Holt and Co., 1938.

RESPONSE STRENGTH IN A CLASSROOM TASK RELATED TO A 'FORWARD' DELAY IN REINFORCEMENT

DONAVON AUBLE AND EDMUND MECH¹

Institute of Educational Research, Indiana University

That a relation exists between degree of learning and delay of reinforcement is well known. However, most experimenters in this area have allowed the 'response' to be made *prior* to the administration of reinforcement. Thus, the gradient which is usually studied is the backward one; this means simply that the response precedes the reinforcement. However, the results of Muenzinger, Dove, and Bernstone (10), Jenkins (7, 8, 9), Thompson and Dove (12), suggest that a gradient of delay also exists in a 'forward' direction; that is, when reinforcement precedes the response.

That a relation exists between amount of learning and immediacy of reinforcement is well known. However, in most studies of this relationship, the experimenter has allowed the 'response' to be made prior to the administration of reinforcement after a specified period of time. Thus, the gradient of delay which is usually studied is the backward one; this means simply that the response precedes the reinforcement. On the other hand, the results of Muenzinger, Dove, and Bernstone (10), Jenkins (7, 8, 9), Thompson and Dove (12), indicate that a gradient of delay appears to exist in a 'forward' direction; that is, with the reinforcement preceding the response.

Since the counterpart of a reinforcing state of affairs preceding the S-R conjunction is apparent in many classroom situations, it

¹ The authors wish to expressly thank both Dr. Wendell W. Wright, Dean, School of Education, Indiana University, and Dr. Nicholas Fattu, Director, Institute of Educational Research, Indiana University, for their support of basic experimentation. Our continuing appreciation is extended, also, to Mr. Homer Terrell, Principal of Central Elementary School, Martinsville, Indiana, who has not only granted permission for a series of classroom experiments, but, in addition, has materially aided in engendering a clearer understanding between psychologist and teacher, a trend, we might add, that could profitably be perpetuated.

appears desirable to extend our notions on this point at the human level.

Hull (5) points out that the significance, if any, of 'forward' delay in reinforcement has not yet been carefully explored, but he is of the opinion that since the reduction of a need necessarily follows rather than precedes the act, it would seem that the 'forward' delay could hardly play an important rôle in selective learning. The problem, then, was to test the prediction that differential performance would result from delayed and immediate reinforcement, when reinforcement precedes the response in a group situation.

PROCEDURE

The subjects were sixty-four third-grade pupils at Central Elementary School in Martinsville, Indiana. Groups I and II were composed of two independent third-grade classes. Day 1 was devoted to obtaining an initial measure of ability for the two groups in a routine computational task. Group I was verbally reinforced on Days 2-5 immediately before being given the task, while Group II was required to wait one and one-half hours after receiving reinforcement before the task was given. On Days 6-8, conditions were reversed. Group II received the task immediately after reinforcement, while the Group I delay between reinforcement and task was one and one-half hours.

The reinforcement was similar to that used by Auble and Mech (1, 2) in previous classroom investigations. The teachers were required to adhere to the following verbal reinforcement throughout the experimental sessions: "You children are doing very nice work on the problems and are improving rapidly. Let me see how much better you can do today. Try to avoid any careless mistakes, and try to do as many problems as you can."

The assumption, of course, is that the verbal stimuli from the teacher, designated as belonging to a class 'reinforcing', will increase the probability that a particular class of response will occur more frequently than if these stimuli were withheld.

The reinforcement was administered at approximately the same time each day. However, the task was either given immediately after, or delayed one and one-half hours in accordance with the operations previously described.

All S's were required to participate in the task as part of the class-

room routine, and, so far as can be determined, were naïve with respect to the fact that an experiment was being carried out.

The task was, in the main, a routine arithmetical one consisting of one-digit number combinations that were to be added or subtracted. Each S was allowed five minutes a day to work the problems, there being enough problems, however, so that the most rapid subject was not able to finish before the time limit. Teachers were cautioned to avoid remedial work dealing with the task throughout the experiment.

RESULTS AND DISCUSSION

Figure 1 shows the cumulated transformed correct responses for Groups I and II under both delay conditions. It should be noted that a square root transformation was applied to the raw data. This was done in accordance with a recommendation by Bartlett (3) because of the proportionality of the successive means and variances, as well as heterogeneity of variance.

An analysis of covariance was carried out between the transformed Day 1 correct responses and the mean cumulated trans-

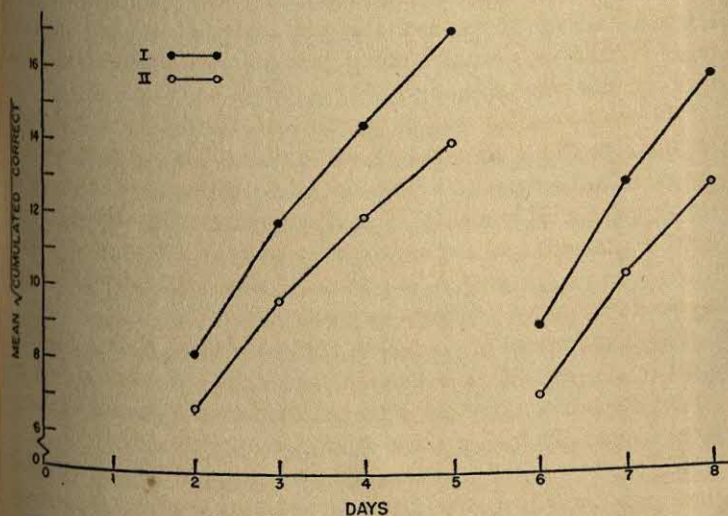


FIG. 1. Mean cumulated transformed correct responses for two groups of S's. On Days 2-5 Group I received no delay, while for Group II a one and one-half hour delay existed between reinforcement and task. Delay conditions for Days 6-8 were reversed for the two groups. (Group N = 32).

formed correct responses on Day 5. The obtained F value of 1.75 was not statistically significant at the .05 level of confidence. The meaning of this insignificant value of F is that it indicates simply that the differences in the cumulated means of the groups of Day 5 can be accounted for by differences in mean level of initial ability as measured in the pre-test trial on Day 1; for as Fisher (4) points out the means of the groups on Day 5 have been 'adjusted' by the analysis to a common mean initial level of performance on Day 1.

Similarly, a covariance analysis was computed between the Day 5 and Day 8 mean cumulated transformed correct responses. This operation was carried out in order to test for differences when delay in reinforcement was reversed from Group II to Group I. The obtained F value of 1.67 was again not statistically significant, indicating that whatever difference existed on Day 8 could not be attributed to the delay in reinforcement. A glance at Figure 1 shows that Group I maintained a relatively superior rank of correct responding over the seven sessions and the delay in reinforcement during Days 6-8 clearly did not exert any suppressing effects.

A similar analysis was carried out for the error responses. Figure 2 shows Group I and II errors under the two conditions of delay for a seven-day period. A covariance analysis between the initial measure and Day 5 indicated that the group difference was not significant ($F < 1$). Finally, the covariance F value of 3.20 that

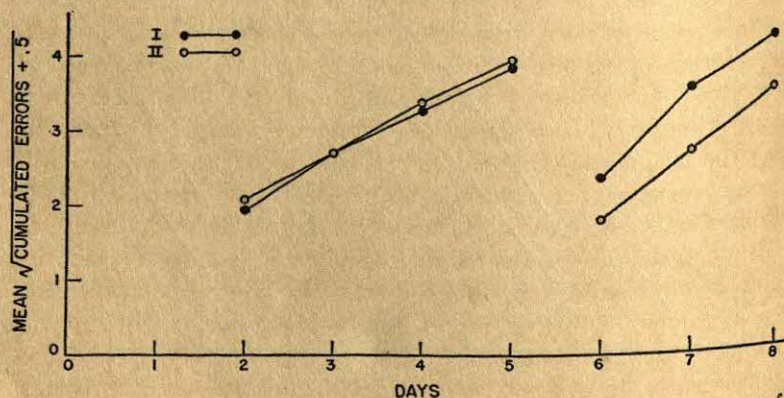


FIG. 2. Mean cumulated transformed error responses for two groups of S's. On Days 2-5 Group I received no delay between reinforcement and task, while for Group II a one and one-half hour delay existed. On Days 6-8 delay conditions were reversed for the groups. (Group $N = 32$).

was computed following the reversal of delay conditions (Days 6-8) also proved insignificant.

Unfortunately no classroom or group data exist, at present, with which to compare the cited results. The data, however, indicate quite clearly no differences existed between the groups that can be attributed to the delay between reinforcement and task. These data are not compatible with the results of previous investigations (10, 12) which utilized infra-human organisms and, especially so when compared with Jenkins (7, 8, 9) whose data strongly suggest that a gradient of delay exists when reinforcement precedes the response. A more specific comparison might be made between Hull's (6) Corollary iii and the obtained data in the present study.

Hull's Corollary iii B states as follows: "The greater the delay in the receipt of the incentive by groups of learning subjects, the weaker will be the resulting learned reaction potential . . ." (6, p. 132).

Since Hull refers to his treatise as a general behavior system, not limited to infra-human organisms, the 'deduction' to be made from the above corollary is that if each subject on each trial undergoes all the delays of the experiment, performance or reaction potential should be weaker as delay is increased. The present empirical data clearly do not correspond to this prediction. Such a result, however, appears to bear out Hull's statement that the corollary in question has yet to be verified in 'detail'.

In a recent review Stroud (11) makes a cogent case for pursuing a program of experimentation to clarify certain notions with respect to a behavior theory for education. Certainly animal experimentation and theory construction have made rapid progress during the last decade, presenting the educational psychologist with many concepts that appear to be testable under classroom conditions. One such concept has been tested in a classroom situation. Although the design was relatively simple, it should be noted that compatible results were obtained from the two independent classroom groups. As further, more intensive classroom experimentation develops and as results are verified by experimenters working independently, it will not be too surprising to witness a revision of many current behavior 'postulates'. There is no reason, then, why a behavior system for education should not basically develop from the crude data and problems of the classroom.

SUMMARY

The present study began with the belief that progress could be made toward evolving a behavior system for education by virtue of testing certain concepts derived from learning theory. Specifically, the question to be answered in the present experiment was: Would differential performance be obtained in a specified task under two varying conditions of delay in verbal reinforcement? The deduction to be made from existing experimental literature on the infra-human level is simply that as delay in reinforcement is increased response strength in a specified task should decrease.

To test whether delayed reinforcement would have less advantage than immediate reinforcement in a classroom situation, the following procedure was used: Two independent third-grade classroom groups composed of thirty-two subjects each worked in a routine task for five minutes a day for eight days. Day 1 was devoted to obtaining an initial measure of ability for the two groups. Group I was verbally reinforced by their teacher on Days 2-5 immediately before being given the task, while Group II was required to wait one and one-half hours after receiving reinforcement before the task was given. On Days 6-8, conditions were reversed. Group II received the task immediately after reinforcement, while the Group I delay between reinforcement and task was one and one-half hours.

The covariance analyses for both error and correct responses indicated that no statistically significant differences existed between groups throughout the experiment that could be attributed to the delay between reinforcement and task. The results clearly were not compatible with a prediction from a theoretical model that was based largely on infra-human data. The study, however, is indicative of a future program of intensive classroom experimentation that if pursued vigorously could contribute materially to the construction of a theoretical learning framework for education.

REFERENCES

- 1) D. Auble and E. V. Mech. "Quantitative studies of verbal reinforcement in classroom situations: I. Differential reinforcement related to the frequency of error and correct responses." *J. comp. Psychol.*, 1953, 35, 307-312.
- 2) —, —. "Quantitative studies of verbal reinforcement in class-

room situations: II. Occurrence of error and correct responses related to a change in stimulus conditions." *J. genet. Psychol.*, in press.

3) M. S. Bartlett. "The use of transformations." *Biometrics*, 1947, 3, 39-52.

4) R. A. Fisher. *Statistical Methods for Research Workers* (6th Ed.). London: Oliver & Boyd, 1936.

5) C. L. Hull. *Principles of Behavior*. New York: Appleton-Century, 1943.

6) —. *A Behavior System*. New Haven: Yale University Press, 1952.

7) W. O. Jenkins. "Studies in spread of effect. I. The bi-directional gradient in the performance of white rats in a linear maze." *J. comp. Psychol.*, 1943 a, 35, 41-56.

8) —. "Studies in spread of effect: II. The effect of increased motivation on the bi-directional gradient." *J. comp. Psychol.*, 1943 b, 35, 57-63.

9) —. "Studies in spread of effect: III. The effect of increased motivation on the bi-directional gradient." *J. comp. Psychol.*, 1943 c, 35, 67-72.

10) K. Muenzinger, C. Dove, and A. Bernstone. "Serial learning: II. The bi-directional goal gradient in the endless maze." *J. genet. Psychol.*, 1937, 50, 229-241.

11) J. B. Stroud. "Educational Psychology." In C. P. Stone (Ed.), *Annual Review of Psychology*, Stanford, Calif.: Annual Reviews, Inc., 1950.

12) M. E. Thompson and C. C. Dove. "A further study on the bi-directional goal gradient in the endless maze." *J. exp. Psychol.*, 1943, 33, 447-456.

BOOK REVIEWS

HELEN M. ROBINSON, Ed. *Clinical Studies in Reading II*. Suppl. Educ. Monog. No. 77. Chicago: University of Chicago Press, 1953, pp. 189. \$3.75.

In this second report from the Reading Clinic of the University of Chicago, emphasis is placed upon visual problems related to difficulties in learning to read. The first part deals with treatment of poor readers with vision problems. Included are discussions of types of problems encountered, visual screening tests used, procedures for referral, and teaching adaptations. It appears that visual difficulties interfere with adequate reading progress in individual cases. The incidence of visual problems among poor readers is sufficiently high to warrant visual testing of all disability cases to identify cases in need of examination by professional refractionists. Results of such examinations provide information useful in remedial teaching.

Four reports of staff research are given in Part II. These reports are concerned with the factors of visual efficiency, personality adjustment, eye-hand preference and reversals, and prediction of increase in reading rate in relation to progress in learning to read. In general, intimate relationship between these factors and reading progress was absent. However, pertinent evaluation of the experimental techniques employed were made with suggestions for further research.

In Part III appear four reports of research by graduate students. Important findings include: (1) A combination of results from visual screening tests plus teacher observation of symptoms provide a better basis of referral to a refractionist than either alone. (2) Girls tend to excel boys in reading through the first three or four grades only. (3) Although auditory acuity and auditory discrimination do not appear to be widespread causes of inefficient word recognition, there is a suggestion that auditory memory span is important. (4) Important factors in leisure reading by retarded readers include desire to read, availability of books of proper interest and easy vocabulary, short books, and a scheduled time for free reading.

Part IV consists of five papers by leaders in the field on vision

and reading delivered at a conference on the subject. Jobe's paper is an excellent discussion of the structure and function of the visual mechanism. Gesell, emphasizes certain developmental factors in vision and reading. The visual problems of poor readers receive adequate attention in an analysis and evaluation by Eames. In her discussion of visual training and reading, Eberl emphasizes the need for a complete case study of all visual functions as well as a visual examination for disabled readers. In certain cases, visual training should supplement remedial teaching to restructure and build the most reliable visual behavior pattern possible. Such training consists of 'teaching to see' rather than just 'muscle exercise'. This approach is not clearly understood by many clinicians.

The editor of this monograph points out that the series of experimental reports and discussions do not provide final answers to the problems considered. However, the contributions are important in advancing our knowledge of the relation of visual defects to reading progress, in evaluation of experimental and measurement techniques, in posing additional problems for investigation, and in pointing out implications of the findings. All workers in remedial reading will profit by careful reading of this monograph.

University of Minnesota

MILES A. TINKER

HAROLD W. BERNARD. *Mental Hygiene for Classroom Teachers*.
New York: McGraw Hill Book Company, 1952, pp. 472.

"This volume is an outgrowth of fifteen years of experience in teaching courses relating mental hygiene to education." (p. vii). The book is in four parts. Part 1, Basic Considerations, discusses the need for the mental hygiene point of view; human needs and mental health; the nature of maladjustment; meeting the needs of children; and special needs of adolescents.

Part 2, Mental Hygiene in the Classroom, deals with teacher personality and pupil behavior; understanding and helping children with problems; the mental hygiene of discipline; teaching methods and pupil adjustment; some questionable school practices; personality problems in the classroom; and constructive classroom approaches to mental health.

Part 3, Special Approaches to Mental Health, includes art as an

approach to mental health; writing for understanding and release; using drama and play; and limitations and precautions regarding mental hygiene.

Part 4, *The Teacher's Mental Health*, discusses a positive view of the teaching profession; the teacher's philosophy—in the school; and the teacher's philosophy—adult mental hygiene.

"Emphasis is placed on the normal child and on mental hygiene in the classroom rather than through the use of clinics, formal guidance services, etc."

"Improvement in educational practices is based on willingness to try theoretically sound methods in spite of beliefs that present techniques are efficacious. But such experimentation should be based on a consistent point of view—there should be a frame of reference. The outlining of such a frame of reference is the task that the writer has undertaken." (p. vii). Apparently mental hygiene is the frame of reference. There are techniques that are only for the use of experts. The writer is concerned in this volume with those which should be known and used by teachers.

This is not a fundamental text in mental hygiene but an attempt to show how mental hygiene can help in the work of the school. It is thus primarily a book on education. Wallin is not mentioned, although Shaffer (1936) is quoted. References are especially to publications concerning teaching methods and teaching aids.

The spread of subject matter can be seen in subjects discussed or references cited: teaching how to write; relations of administrators to teachers; problem of aging; grading systems; need for exercise; learning; projective technique; use of test data; ventilation control; visual aids; therapy through art; sociometry and pupil adjustment; 'experimental approaches' which seem to be 'readily available to ordinary classroom teachers'." Psychodrama and sociodrama (Sociodrama is an extension of psychodrama) (p. 343) are discussed and recommended for the use of the classroom teacher.

The statements, "There is a correct way to rear a child," and "There is a correct way to teach," (pp. 375, 376) lead us to wonder if these have been discovered; but also leave the suspicion that probably there are several ways for each task, and that different ways might be necessary to deal with individual differences. The latter are mentioned, grudgingly, in two paragraphs (p. 146).

The impression given by the book is that many good suggestions

are given for teachers; that special techniques are assumed to be available to teachers but with questionable foundation and training; that there are dogmatic and imperative statements, not all of which can be accepted; some of these are constantly repeated statements found in pedagogical literature, and, though trite, some of them may bear repeating.

The purpose and attitude of the book are good; the style is simple and clear; some terms are used loosely, e. g., experimentation and experimental; the spread of subject matter suggests the survey course which gives orientation but which does not deal thoroughly with any one subject.

A. S. EDWARDS

The University of Georgia

OLIVER L. LACEY. *Statistical Methods in Experimentation*. New York: The Macmillan Company, 1953, pp. 249.

This little text should prove to be an excellent one for a first course in statistics. It touches upon experimental design, probability and the normal distribution, tests of significance of means and of differences between means, Chi-square, correlation and regression, and fiducial limits. Assumptions underlying each statistical technique are explicitly stated and justified verbally. The beginner is not confused by extensive derivation of formulae, an approach which is quite desirable wherever classes include many persons with little or no background in mathematics. Such topics as sampling theory, and the differences between one-tailed and two-tailed tests of significance, for example, are introduced easily and simply in a way which should provide a good foundation for later expansion. The clarity of exposition is remarkably good. One might wish to complement the discussion of small sample theory, and perhaps applications of the Chi-square test, if one's class moved fast enough to permit it.

Many examples and problems are included. Usually these are at the ends of chapters, but some are within the chapters, close to pertinent discussion. These are in addition to the illustrative examples used by Lacey in his discussion. The examples are to be worked by the student, and there are answers given for these, along with some very good explanatory material for each item. Problems follow, involving the operations just discussed in the

chapter and illustrated and explained in the examples. For the convenience of the instructor, answers ought perhaps to be provided somewhere for these also, although they may be easily worked out in a short time. The examples and problems are nicely distributed over several kinds of research studies, so that applications of given techniques to various fields are well illustrated.

Several tables are included, not all in the form usually encountered in similar texts. Table B, for example, is an abbreviated version of the familiar table of percentages of total area under the normal curve between the mean ordinate and other ordinates. It would require the student to interpolate more frequently than with the fuller table, which might provide an additional hazard for the beginner.

As one who has taught elementary statistics for several years to students who have almost no background for the course, the reviewer believes this book comes as close as any he has seen to being an ideal teaching text for that kind of class. It may be considered too limited in scope by those who have more advanced classes, or who wish their courses to cover more ground. For students in psychology, education, sociology, and related fields who are limited in background, however, this text should provide an excellent introduction to statistics and experimental design which involves very little suffering.

CECIL M. FREEBURNE

Bowling Green (Ohio) State University

CHARLES KOCH. *The Tree Test*. New York: Grune and Stratton, 1952, pp. 87.

In this book are presented the interpretations which the author has found applicable to elements of the form, location, and graphological features of drawings of trees, used as a projective technique. For this purpose the subject is requested 'to draw a fruit tree'. The author further suggests that the subject draw additional trees "if one has the impression that the first drawing obtained does not correspond with the person tested or that the drawing differentiates too little." He claims that such successive productions tap different layers of the psyche.

The introduction, which comprises the first half of the book, consists of a lyrical and imaginative treatise on the symbolism of

the tree followed by several illustrative analyses of test subjects. With regard to the validity of his interpretations the author states only that they have been proved correct "(1) by the nature of the recorded expression, (2) by comparative investigations, by direct observations, by testing and by judgements of independent persons." The test has been employed mainly in the vocational guidance of adolescents and adults, and the author points out with proper caution that additional samples would be needed to extend the interpretation of tree drawings to special groups such as the mentally deficient or the mentally ill. He also suggests that the test be used in conjunction with other tests for diagnosis and counseling so that the findings may check and augment each other. It is unfortunate that the author does not see fit to extend the same cautious and critical attitude toward the interpretations themselves.

The only objective data contained in the book are tables presenting the developmental trends for various tree-drawing characteristics in the age range from five to sixteen years, and a comparison between two groups of the same age but different mental endowment. Unfortunately, all statistics are presented as percentages with no indication of the sizes of the samples studied.

The second half of the book contains clear and detailed illustrations of possible variations in the drawings of root, trunk, and crown together with lists of interpretations for each detail. The many different interpretations suggested for any one drawing characteristic should give the psychodiagnostician ample leeway for employing his imagination in producing a rich character portrayal of the subject. However, being a skeptic with regard to the principles of graphological interpretation and the "scientific analysis of forms of expression," this reviewer is inclined to wish for less fantasy and more fact.

This book may have some value for those who are particularly interested in drawing as a projective technique. It would be quite interesting to compare the interpretations given here with those offered for the tree drawings on the HTP test. It might also be of interest to determine whether the same developmental trends obtain in this country as those found by the author for Swiss children.

GOLDINE C. GLESER

Washington University School of Medicine

JAMES L. MURSELL. *Using Your Mind Effectively*. New York: McGraw-Hill Book Company, 1951, pp. 254.

In spite of the more general implication of the title and the author's protestation against confining the concept of mental effectiveness to scholastic efficiency, *Using Your Mind Effectively* can be described best as a 'how-to-study' book written in popular style. The theme upheld throughout the volume is that the achievement of better thinking is the requisite for attaining greater adeptness in studying. Since the desideratum of better thinking is learnable and teachable, Mursell's task is to show the way to the reader. The steps in reaching comprehension and understanding are formulated as: (1) obtaining an over-all view of the material under consideration, (2) identifying its main features, and (3) working the details into their proper relationship to the main skein.

After an introductory section on the value of mental efficiency, the book is broken down into three main divisions; one dealing with an exposition of the general psychological principles to be observed in attaining intellectual mastery and two devoted to the demonstration of practical applications. In Part One the author elaborates on the nature and usefulness of the sequence; picking up the main thread, identifying essentials, and relating the details to the whole. He discusses the importance of this sequence in the reading of textbooks, study of college courses, and the writing of term papers and theses. The construction of a mental map into which essential parts and details may be fixed is offered as the *modus operandi* for achieving mental excellence. Applications of this technique to non-academic situations are demonstrated, e.g., preparing a talk, orienting oneself to a new job, arranging one's budget, and learning to play tennis.

In Part Two standard topics of the 'how-to-study' books such as budgeting time, note-taking and note-using, self-testing, and concentration are discussed under the rubric of working tools and practical plans. The general rationale for the recommended approaches as well as the minutiae of procedural detail is set forth. Much of this is the stock advice on these topics of 'how-to-study' books. The author appears at times to be self-conscious about some of the trivia which he presents. After marshalling a lengthy list of materials and equipment that should be kept in the study room, he feels constrained to remark that he offers this for whatever it is

worth. In this division of the book the *leitmotif* is that the facilitation of thinking should be the governing consideration in the arrangement of time, conditions, and methods of work. Everything should be instrumental to the goal of encouraging thinking; note-taking and review are seen as experiences in thinking.

In Part Three under the title "Some More Extended Applications" appear the following topics: memorizing, reading, writing term papers and theses, and creative thinking. Good memorization is viewed as conditional upon proper purposing, understanding, and noticing. Analyzing, interpreting, and thinking are considered the crux of the development of reading skill. In writing term papers, it is recommended that the outline and over-all plan be completed before a single word of the final manuscript is written. In a very short chapter on creative thinking, Mursell makes the point that whenever one grasps the meaning of another person's writing, one is engaged in creative thinking. The author makes clear that his aim in offering suggestions is not to facilitate the functioning of routine-minded workers, but to influence the intellectual worker to approach even mundane tasks as invitations to creative thinking. He concludes his work with the statement that he has tried to make everything in the book center on creative thinking.

After finishing this book, the reader might legitimately ask what progress psychology has made, since William James, in offering assistance to students in their pursuit of academic success. No later writer seems able to go beyond or even approach the level of James' *Talks to Teachers on Psychology: and to Students on Some of Life's Ideals* published in 1899. Basically little that is new since Watson's *How To Use Your Mind* appears evident in the more recent books on 'how to study'. Aside from the Q 3 R technique of Ohio State University for studying chapters in the conventionally organized textbook and the 'push on' or quick reading method for mastering foreign languages, there is in the book under review scarcely any preference to new developments in effective study methods. Psychologists who choose to write 'how-to-study' books seem to feel obliged to approach their topics as exclusively cognitive-intellective matters unrelated to the emotions and the total personality. They appear to be oblivious to the psychoanalytic movement and to dynamic psychology in general. Mursell fails to suggest that the inability to understand and think effectively may be due to emotional factors as much as to poor techniques in

cognition. The failure to integrate cognitive functions with the entirety of the personality accounts for the antediluvian character of most 'how-to-study' books.

Mursell has done a good job of expounding the traditional approach for developing mental efficiency. However, at times, the style is repetitious, long-winded, and too general without the compensation of being inspirational—hence boring. The standard of writing at times seems to dip below the college level to which it is apparently directed. The attempt to offer applications to non-academic pursuits appears to be simply a gesture. For those who want a simple and very clear exposition of the conventional views on effective use of the mind, this book may be most satisfying.

PHILIP M. KITAY

Adelphi College

C. W. ODELL, *How To Improve Classroom Testing*. Dubuque, Iowa: William C. Brown Company, 1953, pp. 156.

Professor Odell has prepared a student manual that emphasizes the practical and non-technical aspects of the construction and administration of informal or teacher-made tests of achievement. Students of education will be pleased to find that he sets the problem of measuring achievement in the context of curriculum development by taking the objectives of education as the definitions of the achievements that are desired. He then concentrates on testing programs and types of tests, excluding problems of intelligence and personality measurement. Chapters IV through XII are devoted to test construction principles and illustrations of various types of test items or problems; these illustrations are drawn from a number of subject matter fields. The manual appears to be a good source of ideas for test item types, and as such it should be of value to teachers. One chapter on administration and scoring and one on statistical methods in connection with testing complete the volume. The chief contribution of the manual is its rather extensive and practical advice on how to develop and use informal tests.

CHESTER W. HARRIS

The University of Wisconsin

PHYLLIS GREENACRE. *Trauma, Growth and Personality*. New York: W. W. Norton and Co., 1952, pp. 328.

In the past decade Greenacre has published many psychoanalytically oriented studies of early personality development. These have appeared in many journals; most of them in the *American Journal of Orthopsychiatry* or the *Psychoanalytic Quarterly*. The main purpose of this volume is to make a selection from these readily accessible to more readers. The connecting link for the selection included is that they are all related to the interaction of maturation phases and special traumas in the first few years of

Studies included in this volume are The Biological Economy of Birth; The Predisposition to Anxiety; Infant Reactions to Resonant; Urination and Weeping; Vision, Headache, and the Halo; Anatomical Structure and Superego Development; Conscience in the Psychopath; A Study of Screen Memories; The Prepuberty Trauma in Girls; Problems of Early Female Sexual Development; Respiratory Incorporation and the Phallic Phase; and Some Factors Producing Different Types of Genital and Pregenital Organization. The volume contains a very favorable and informative introduction by Dr. Ernst Kris, a list of references for each chapter in the back end of the volume, and an index.

The studies will meet the unreserved approval of that school of psychoanalysts and clinicians who have an interest in and include in their studies social as well as individual factors. This group will probably react to the content of the volume and the interpretation pretty much the same way as Kris does in his introduction. Kris sees Greenacre's studies as a valid elaboration of Freud's concept of the trauma which enables Greenacre to enlarge her vista beyond the range of psychological events and to reemphasize the continuity that links earliest physiological reactions of the organism to its gradually differentiating psychological experiences. Like him, too, they probably would consider the volume as composed of studies which are evidence of Greenacre's "happy and productive fusion of observational and speculative writing." Experimental psychologists obviously would be less satisfied. In these studies they will find no controlled investigations and nothing treated statistically. They will instead find clinical and observational reporting. However, in all fairness they will have to admit that Greenacre, though

primarily an analytic clinician, shows a clear familiarity with and knowledge of contributions from related fields as well as psychoanalysis. For example, on the physiological level she shows familiarity with the works of Cannon, Morgan, and Coghill. On the psychological level she shows familiarity with the writings of such contributors as Gesell and Halverson, Mandell Sherman, Sontag, and Jersild.

H. MELTZER

*Psychological Service Center
St. Louis, Missouri*

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 45

April, 1954

Number 4

AN EXPERIMENTAL COMPARISON OF RECI- TATION, DISCUSSION, AND TUTORIAL METHODS IN COLLEGE TEACHING¹

HAROLD GUETZKOW

Carnegie Institute of Technology

E. LOWELL KELLY and W. J. McKEACHIE

University of Michigan

In 1946-47 the Psychology Department of the University of Michigan instituted a seminar with academic credit for graduate students who were teaching sections of the introductory course. During the first semester the seminar reviewed the literature on experiments in teaching methods. They then wanted to evaluate their own teaching techniques. This experiment was undertaken as a 'work shop' project of the seminar.

During the second semester of 1946-47 there were 865 students enrolled in the elementary general psychology course at the University of Michigan. The course carried three hours' credit. Each student spent one hour a week in a large lecture section and two hours in small class sections. The students were divided into three lecture groups ($N = 250$ to 300) and twenty-six class sections ($N = 25$ to 35). The same lecture was delivered three times on Mondays and Tuesdays. The two weekly class meetings were held on Wednesdays and Fridays. The lectures were given by a senior staff member

¹ The following assisted in planning and carrying out this experiment: Robert D. Boyd, Everett W. Bovard, Robert G. Givvy, Theodore N. Harrison, William H. Hossick, Norton B. Knight, Ernest J. London, Elizabeth A. Malcolm, Herbert H. Meyer, Elliot G. Mishler, Harvey Schrier, Audrey A. Snyder, and Philip I. Sperling.

and the small section meetings were conducted by psychology graduate students.

THE PRIMARY FOCUS OF THE EXPERIMENT

In common with others who take seriously the task of teaching, the seminar group often found its deliberations ending in disagreement. We concluded that the primary prerequisite for our work was the construction of better instruments to evaluate the outcome of our teaching in elementary general psychology. It was assumed that the sensitivity of such instruments might be determined by whether they discriminate between teachers of varying quality and by whether they reflect differences in teaching procedure. It was important to develop external criteria of the efficacy of our teachers and to teach sections according to widely divergent methods.

In order to determine what evaluative measures were needed, it was necessary to decide upon the objectives of the general psychology course. To aid in the definition of these objectives each member of the seminar presented his view as to what the objectives of the course should be. Two student sections in the general psychology course also discussed goals of the course. Results of these discussions and papers were then considered and the following tentative goals were formulated:

- 1) Understanding of the facts and principles of psychology basic for further study of psychology or for use in other areas of life.
- 2) Ability to apply learning to aid the student's understanding of human behavior in his everyday life.
- 3) Development of a scientific and analytic attitude toward human behavior: (a) Understanding multi-causality. (b) Familiarity with the experimental approach. (c) Increased skill in formulating conclusions. (d) Increased skill in evaluation of conclusions. (e) Ability to distinguish between fact and opinion.
- 4) Awareness of the adaptability of human nature. (a) Recognition of the limitations of attempts to change adult human nature. (b) An attitude of understanding toward individual differences.
- 5) An appreciation of the value of psychology.

THE DESIGN OF THE EXPERIMENT

The design of the experiment was dictated by a number of conditions. First, of course, was the imperative consideration that we needed to teach the elementary course in such a way that all stu-

dents passing the course, regardless of the method used, were qualified to go on to the other courses in psychology offered by the department. Second, the fact that we had multiple objectives demanded we use a wide variety of criterion instruments. A number of instruments died a-borning; sometimes their reliabilities were simply too inadequate. In the end some thirteen instruments were utilized to provide criterion measures. Third, we noticed the relative inability of previous researchers in the area of teaching to secure differences, no matter what instruments were used. We wanted to contrast teaching methods which at least on a priori grounds might be differentially effective. In addition, we wanted to test some of the differences in method we were hotly arguing about in the seminar from time to time. It was practical to compare three methods. In summary, the design may be stated in the null form: There will be no significant differences in outcome with respect to course objectives between students whose class meetings are taught by Method 1 and those students who are taught by Method 2 or by Method 3.

THE TEACHING PROCEDURES

Eight teaching fellows each taught three sections—one by each of the experimental methods. Thus eight of the twenty-four sections² were taught by each of the three methods. In addition to the common lecture, all students were given the same assignments, used the same textbook, and were provided with the same opportunity to see educational movies and take sample tests, as outlined in the syllabus. Hence, the students were given a common core of lectures, movies and test demonstrations, and text and supplementary reading assignments. The teaching method was varied only with respect to the procedures used in the two section meetings per week. No explicit announcement was made to the students that an experiment was in progress.

During the week before classes actually started the following memorandum was issued and discussed by the eight instructors who were teaching the twenty-four sections of the elementary course:

The suggested variable in the experiment for the spring semester is teaching technique. Three types of teaching technique would be

²Two sections taught by the course administrator were not used in the experiment.

compared: Drill—recitation; group—discussion; and study—tutorial. This variable was chosen rather than class size because experimental evidence indicates that variation in size produces negligible variation in performance, unless technique is also varied. These three techniques represent different degrees of directed vs. non-directed and group vs. individual work.

1) *Recitation—Drill.*—This method will be instructor directed, instructor paced, instructor dominated. A quiz will be given at each class meeting. The instructor will bring a list of prepared questions to class. Students will be called upon to answer and will be marked for recitation. Each day the instructor will present a short lecture-drill on the fundamentals.

2) *Group—Discussion.*—The discussion section is to be regarded as a supplementary tool used in conjunction with the text, the lectures, and the readings. It operates on the assumption that merely telling the student or having him read is insufficient. Students must formulate and discover the relationships themselves. Instead of drilling facts, it attempts to clear up misconceptions and integrate the concepts of the course.

The discussion group creates an atmosphere in which the student feels free to expose his misconceptions, so they can be corrected; he is encouraged to bring in the problems which have troubled him in the assignment; he is stimulated to utilize the basic concepts of the course to interpret incidents which occur in his day-to-day living. An attempt should be made to establish the sense that the class works together; each person learns from the other as well as from the instructor. Being a member of a group, he is carried along by the group's enthusiasm and interest. The student should be given to feel that only by actually participating in the discussion will he, himself, gain a usable understanding of the content of the course.

Functions of the instructor.—At various times the instructor will serve in the following capacities: Chairman: presiding, deciding order of speakers, etc.; Summarizer: at appropriate times throughout and at the end of each class meeting, he should summarize the ideas discussed. Sometimes this summarizing function will express itself as a resolution of a conflict among discussants. Each class meeting should end with a sense of having actually covered ground. Discussion Stimulator: on needed occasions the instructor will ask inference questions, read case-histories, perform demonstrations,

etc., to stimulate fruitful controversy and increase the 'reality' of the discussion. Informant: at times the instructor will supply data and expert knowledge, because of his background in psychology. At no time will the instructor lecture. If he desires to present material not included in the readings or lecture, he should ask a number of the better students to prepare a symposium on the topic, or use some other student-participant device.

Special Facilitating Devices.—In order to facilitate the discussion atmosphere, the following techniques are suggested: Techniques to reduce the restraint: (1) no grade should be given on the discussion itself; this may allow the student more freedom in revealing his ignorance and misconceptions; (2) by arranging the seats in a U shape, it may be possible to remove part of the traditional classroom restraint; the students then can talk to each other as well as to the instructor; (3) have the students introduce themselves on the first day; supply the class with a seating chart and names of fellow-classmates; use first or only last names (omitting Miss or Mr.) to create informality.

Techniques to spread participation: (1) encourage comments made by reticent students; quash 'stars' who monopolize; (2) after a few weeks, ask backward students to bring written questions, which can be read aloud. Sometimes have all the students answer a question raised by writing out the answer during classtime; this gives the poorer students a chance to participate. Have students look up facts and answers in their text, so that poorer ones can read answers and learn to contribute, using text as prop; (3) essay thought-type tests, once every other week, except when course exams are given.

Discussion instructors might introduce their work by mentioning Lewin's experiments on group decision.

3) *Tutorial—Study.*—This method will emphasize an individualized approach with the student pacing his own work within the framework of the course. No quizzes will be given, but occasional written papers will be required. The class work will be relatively undirected as compared with the Drill—Recitation method. The instructor will bring other elementary texts for references and demonstration apparatus to class. He will not conduct the class formally, but will be available to the students for consultation throughout the hour. The students will utilize the hour for study and consultation. The instructor will endeavor to help the student

to find the answer to his questions rather than answering them himself. Social conversation will be prohibited. This method is based on that reported favorably by Scheidemann (*Sch. and Soc.*, 1927, 25, 672-674.).

INSTRUCTOR VARIABILITY

In comparing teaching methods, one is always faced by the problem of the interaction between teacher and method. We chose to assign each instructor to teach three sections—one by each of the three methods. To reduce differences in instructors' abilities and motivations for teaching the different methods, the entire group actively participated in the general planning for each teaching method. In addition, three special committees were appointed, one to represent each of the three methods. Each committee was responsible for seeing that its method was used effectively. Thus, the 'recitation committee' worked out quiz questions and other aids for the Recitation—Drill classes; the 'discussion committee' provided lists of suggested discussion questions for each day's Group—Discussion classes, and the 'tutorial committee' provided demonstrations and reference books to be taken to the Tutorial—Study classes. We attempted in this way to make sure that each method got a fair trial.

The time sequence of methods was varied from instructor to instructor so that no method would consistently be used for the instructors' first sections to be taught during the week. All permutations of the order of methods were used.

EQUATING OF GROUPS

While we did not know what factors enter into students' choice of sections, we suspected that we could not assume that our classes were equated in intelligence and other possibly relevant variables. During the first and second class periods of the initial week of the semester, basic sampling data were obtained and an Otis Test of Mental Ability (Higher Form A) was administered to each section.

To carry out the process of equating groups, summary sheets for each section were compiled. For all hours, with two exceptions, all teaching techniques were represented each hour. It would have been inconvenient to attempt to change the time-schedule of the students. Hence, students were shifted to equate sections within each hour. Although it was not possible to obtain complete equa-

tion of the sample for all sections on all characteristics, we were able to make the students of each section typical, in the sense that each had an appreciable number of women, veterans, etc., and mean score intelligence and grade point average were perfectly matched.

During the first period of the second week, the students were asked to give order of their preference for the three teaching methods: Recitation, Discussion, Tutorial. This was treated lightly, with the suggestion that there was much disagreement among the students about the best method. After the preferences were collected, the instructor presented his rationalization for the method to be used in this particular section. We wanted to help the students feel that they should give the method a try; that even though it might be new to them, it had been used successfully before. If a student asked whether an experiment was being conducted in the class, he was told that because no crucial experiments on teaching had been done, various methods were always being tried out by different teachers.

The students were asked to indicate their method-preferences again at the end of the semester. The instructors also indicated their preferences for teaching method at both times.

MEASURES OF OUTCOME³

Course examinations.—Academic achievement in the course was judged primarily on the basis of scores on objective achievement tests. Three course-wide objective tests were given. The first took place at the end of the initial four weeks of the course. A mid-term examination was given after eight weeks of the course, and a final examination at the completion of the course. Coverage in these examinations was retroactive, in that each sampled all of the material covered up to that time in the course. During the semester only standard multiple-choice questions of the one-best-answer type were used.

The system used for constructing the examinations was as follows: H. H. Meyer was placed in charge of the construction of the

³ To save printing costs, detailed tables of results and all measures used in the experiment are included in an appendix which has been deposited with the American Documentation Institute. Order Document No. 4033 from American Documentation Institute, 1719 N. St., N. W., Washington 6, D. C., remitting \$2.00 for 35 mm. microfilm or \$3.75 for 6 x 8" photocopies.

examinations. The other teaching fellows submitted questions and suggestions on the respective assignments. On the basis of the course objectives and the material covered, Meyer drew up a table of specifications for the examination to be given, selected items submitted, and wrote new items to fit the table. This preliminary edition of the list was submitted at a meeting of all teaching fellows, and items were altered or eliminated. The revised examination was then further edited by Meyer, and the final edition constructed.

Since one of the objectives of the course was that the students should be able to apply to life-situations the principles they had learned, many of the items were 'application' questions. The question then arose as to whether such questions were actually measuring the intelligence of the students more than their knowledge of psychology. This hypothesis was tested by dividing the final examination into two parts, consisting of fifty questions testing the memory of facts and fifty questions of the application type. The scores of these parts were correlated with the students' scores on the Otis examination, which was used as a measure of intelligence. The hypothesis was sustained to some extent. The correlation between Otis and Factual question scores was only .17, while the correlation of Otis and Application question scores was .41. The *t*-test of the difference between the correlations is 1.6—significant at the .05 level.

USAFI Examination.—In addition to the tests and final examination, the United States Army Forces Institute Examination in elementary psychology was used as a measure of achievement. In 1946-47 this test was one of the few standardized achievement measures in general psychology.

We had also hoped to use essay tests as measures, and spent a good deal of time attempting to develop reliable scoring methods. However, we were unable to raise the scorer-reliability above .60 on the kind of broad questions which seemed necessary to us if we were to develop measures of anything more than the achievement sampled by our objective tests. Hence, essay tests were not included in our battery of measures.

Misconceptions.—One of the objectives of the course was to teach in such a way that the student's knowledge of psychology was carried over to his everyday thinking. If this were accomplished, common misconceptions about human behavior should be effected

even when the misconceptions are not specifically discussed. Hence, a test to measure this outcome was devised by McKeachie.

This test, consisting of three parts, was constructed from common misconceptions submitted by the teaching fellows and from Valentine's list (7). Part 1 consisted of eleven multiple-choice items. Part 2 consisted of seventeen true-false items, the conventional form for misconception items. Part 3 consisted of ten proverbs. These statements were common stereotypes meeting acceptance in everyday conversation, which from a psychological viewpoint were not completely accurate. The students were asked to decide whether the proverb contained more truth or falsehood. The split-half reliability of this test was .73 (estimated for the whole test).

Conceptions of parenthood tests.—Duvall (2) has classified responses of four hundred and thirty-three Chicago mothers of different class levels to the two questions: "What are five things a good mother does?" and "What are five things a good child does?" into two categories—'traditional' and 'developmental'. The traditional responses are characterized by a desire to make children conform to a rigid conception of the rôles of mother and child. The developmental responses are characterized by respect for the child as a person and by expectations of mother and child in terms of growth and development rather than as specific behavioral conformities. A course in general psychology might be expected to produce a more developmental conception of the rôles of mother and child. With this hypothesis a test was created by McKeachie using responses listed by Duvall as typical of traditional and developmental conceptions. Split-half reliability of the test was .80, estimated for the whole test.

Scientific and analytic attitude toward human behavior.—McCandless' (5) test of analytic attitude toward human behavior was our measure in this area.

To measure our objective of creating a favorable attitude toward psychology an eight-item Likert-type scale was constructed. The split-half reliability of the test was .75. As another indication of the attitude of the students toward psychology, students were asked to check the advanced courses in psychology which they would like to take on a list of those offered to undergraduate non-majors. A potentially significant measure of the student estimate of value of the elementary psychology course is the number who decide to

TABLE 1.³—MEAN SCORE SUMMARY OF RESULTS

Measure	Recitation- Drill Mean Score	Group- Discussion Mean Score	Tutorial- Study Mean Score	CR for Differ- ence Between Means of Highest and Lowest
Final Examination	97.7	96.5	95.8	2.7
USAFI	57.2	57.4	57.3	*
Misconceptions	49.6	49.3	48.9	*
Parenthood	8.9	8.6	8.8	*
Analytic Approach	7.6	7.6	7.8	*
Attitude Toward Psychology	21.3	21.9	20.8	1.8
Number of People Planning to Con- centrate	19	7	11	2.4
Number of Advanced Courses Stu- dent wants to take	4.3	3.9	3.6	2.9

* Less than 1.0.

concentrate in psychology while enrolled in the course. Students were asked at the beginning and end of the course in what subject they intended to concentrate.

RESULTS IN AREA OF ACADEMIC ACHIEVEMENT

The traditional method of measuring course outcome is a final examination. One of the students' chief goals in a course is to pass the examination. As indicated in Table 1, with respect to this conventional criterion our three methods ranked in the order: Recitation—Drill; Group—Discussion; Tutorial—Study. It should be noted that the absolute difference between the means of Recitation and Tutorial method is only 1.9 score points, and, although statistically significant, it is so small as to be of dubious social significance.

Not only are the differences between instructors not statistically significant, but there was no significant interaction between instructor and method. These results bring small comfort to those who look upon teaching as an art dependent upon the personality of the instructor.

What was the relative effectiveness of each method for different degrees of intelligence within the range of this class? Measured by scores on the final examination, the recitation method was found to be more effective for all levels of intelligence in our group. This was contrary to our general expectation that the recitation method would be effective for poor students, but not for bright ones.

The United States Armed Forces Institute Examination in elementary psychology (Form A) was given to about two-thirds of the students enrolled. All of the students in designated sections took the test so that there was no selective factor. Students were informed that the test would not count in their grades, and they need not make special preparation for the test. They were urged to do their best, but inasmuch as they knew their score would not count in a grade, many students were probably not as highly motivated as if it had been their final examination. In spite of this lack of motivation, a mean score for all three groups was about the 70th percentile on the norms given for the test. Differences between means for groups taught by different methods were not significant.

The possibility that one method stimulated or forced students to do more work outside of class than the other methods used was considered. The students were asked during the fourteenth week to estimate, anonymously, the total time they had spent on psychology outside of class in a two-week period when there had been no course-wide tests. Results showed negligible differences between the three methods. This measure may be inappropriate, in that a good teaching method may stimulate a student to more study, while a poor one may so confuse him that he is forced to spend more time in study.

OUTCOME IN THE APPLICATION OF PSYCHOLOGY TO EVERYDAY LIVING

Like Horrocks (4) we find nothing in our results to encourage hope that one of our methods is superior in teaching students to apply psychology to life. There were no significant differences between method on either the misconceptions test or the parenthood test. However, like others who have constructed misconceptions tests, we found that college students accept a number of common misconceptions. Every misconception in our test was accepted as true by at least some students. And this after the elementary semester course!

There were no significant differences between methods measured by student scores on the McCandless test.⁴

⁴ Such negative results may be due to insensitive measures of ineffective variables. A pre-course post-course comparison of scores on these measures was obtained later. There were no significant gains on either the misconceptions or the analytic thinking tests.

Results of our scale of attitude-toward-psychology showed students in the group—discussion sections to possess slightly more favorable attitudes toward psychology than students in tutorial—study sections ($P = <.03$), with students in recitation not significantly different from either of the others. However, most students in all sections had distinctly positive attitudes toward psychology as measured by our scale.

Students being taught by different methods differed in the number of advanced psychology courses which they wished to take, with Recitation—Drill again leading. The critical ratio for the difference of the means of the recitation and tutorial sections is 2.9, which is statistically significant. A follow-up study three years later indicated that students in recitation sections actually did take more advanced psychology courses than students in other sections. Note that in the class as a whole, regardless of method, the average student would like to take four more psychology courses. Frankly, we were somewhat surprised by this interest in further course work in psychology.

While the numbers involved are small, a significantly larger number of students in recitation sections chose to concentrate in psychology than students in the other two types of sections.

As another measure of the value attached to the course we had hoped to use the number of cancelled registrations. However, exclusive of those who dropped the course during the first week, i. e., before the different techniques of instruction were introduced, only seven students dropped out of the course. Three of these students were in discussion sections and two each in recitation and tutorial sections. Contrary to the prophecies of some of the local Jeremiahs, there was no mass exodus from any method.

ATTITUDES OF STUDENTS AND INSTRUCTORS TOWARD TEACHING METHODS

An indication of the students' attitude toward the method of instruction used is the shift in their preferences in the beginning of the experiment to their preference at the end of the course. At the beginning of the semester the recitation and discussion methods were approximately equally preferred and the tutorial somewhat less so. Students in recitation classes showed a statistically significant gain in their preference for recitation, while students in dis-

cussion and tutorial sections showed no significant changes in preference.

What is the relationship between the student's preference for a method and his success in learning by that method, as measured by the final content examination? It will be remembered that students originally were matched on the basis of grade-point average and Otis scores within each method. Using scores on the final examination as a criterion, we found no significant differences between students who initially preferred the method used and those who initially ranked as least preferred the method to which they were actually subjected.

The attitudes of the instructors toward the three methods of teaching were assessed at the beginning of the semester and again at the end of the sixteenth week of the semester, before any of the instructors had knowledge of final results. At the beginning of the semester, there was a slight preference for the group discussion method, with recitation-drill and study-tutorial methods preferred about equally. By the end of the semester preference for the recitation method had increased to approximately equal that for discussion while preference for tutorial had dropped appreciably. No teaching fellow ranked the tutorial method higher at the end of the experiment than he had in the beginning. It is interesting to note that while some instructors still preferred teaching by the discussion method and expected results to justify their preference at the end of the semester, their preference was not borne out by the results. Yet most of our teaching is based upon what methods teachers prefer.

DISCUSSION

By and large, we found no differences between the three teaching methods from the point of view of educational outcome. The few statistical differences, in general, favored the recitation-drill method. The results of the experiment, of course, were contrary to our original expectations. They constitute a clear confirmation of the general conclusion derived from experiments on instructional procedures since the early 20's. As Good recently put it, after reviewing the literature in this area, "The complexity of the teaching-learning process is such that attempts to establish the relative merit of a 'general method' of teaching are likely to prove incon-

clusive" (3, p. 275). Our results are thus in line with other investigations, despite the fact that our teaching methods seem to be much more extreme in their differences than had been the case in earlier experiments.

One of the criticisms which is commonly made of the earlier experiments on instructional procedures is the limitation of the appraisal instruments, inasmuch as they usually measured only memory-knowledge outcomes of the particular course. Our experiment included a number of measures of attitude. Yet we obtained few differences among our groups with respect to terminal attitudes in the course. Perhaps the 'general methods' comparison is also going to prove unfruitful even when a broader set of educational outcomes are evaluated.

Our experiment is inadequate in two important ways. Little work was done using delayed measures, which might have turned up differences between our methods after six months or more had elapsed after the end of the course. Perhaps an even more significant inadequacy of our experimental design rests in the fact that the instruction was given by inexperienced teachers who were just at the beginning of their college teaching careers. Cogent to this limitation is Peters' (6) experience that as his high-school teachers continued their work with a new technique into the second year, the difference between his traditional and experimental groups became stronger, as exhibited on a test of information.

Perhaps the most important factor which tends to eliminate differences in comparing different teaching methods is the fact that the experimental manipulation of the instructional procedure affected such a small segment of the students' learning process. The student comes to us from a long series of classroom experiences which have emphasized content acquisition. His three semester hours in the elementary psychology course are only three in the 120 semester hours he needs for his graduation from a large university. Even within our course we gave common examinations to all of our students, regardless of the teaching methods used. They attended the same lecture each week. They all had opportunity to see the same movies. With well-matched student groups and a design which held the personal characteristics of the instructors constant, we obtained no differences among our students when we subjected them to comparatively small differences in environment. There was more difference between our teaching methods than in

most other experiments reported to date. We used a wider variety of outcome measures than had been used up to 1946. Yet, the variations in our instructional procedures were too small, when contrasted with the other factors which remain constant, to greatly affect the outcomes of the learning process.

We still believe it is of vital importance to gain a better understanding of the effects of differences in instructional procedures. But the fact that the classroom situation involves so many variables, and that the interactions among these variables are considerable, means that more fundamental research is needed upon the quasi-isolated sub-systems within the total educational process.

REFERENCES

- 1) C. E. Buxton. "The pros and cons of training for college teachers of psychology." *Amer. Psychologist*, 1949, 4, 414-417.
- 2) Evelyn Duvall. "Conceptions of parenthood." *Amer. J. Sociol.*, 1946, 52, 192-203.
- 3) C. V. Good. "Colleges and universities: VIII. Methods of teaching." In W. S. Monroe (Ed.), *Encyclopedia of Educational Research.*, pp. 273-278. New York: Macmillan, 1952.
- 4) J. E. Horrocks. "The relationship between knowledge of human development and ability to use such knowledge." *J. appl. Psychol.*, 1946, 30, 501-508.
- 5) B. R. McCandless. *McCandless Attitude Toward People's Behavior Test*. Mimeographed material available from author.
- 6) C. C. Peters. *Teaching High School History and Social Studies for Citizenship Training*. Coral Gables, Fla.: Univ. of Miami Press, 1948.
- 7) W. L. Valentine. "Common misconceptions of college students." *J. appl. Psychol.*, 1936, 20, 633.

VALIDITY OF AN OBJECTIVE EXAMINATION FOR THE PLACEMENT OF FOREIGN STUDENTS IN ENGLISH COURSES

IRVING LORGE

Teachers College, Columbia University

and

LORRAINE KRUGLOV DIAMOND

The City College, New York

With the post-war influx of foreign students to American universities the problem of adjusting the college curriculum to the needs of foreign students became a pressing one. The need to place foreign students in English courses at levels appropriate to their mastery of the English language has been one of the major factors leading to revisions in the curricular offerings for foreign students. In addition, it forced the exploration for techniques by which such students could be placed in appropriate English courses.

A common-sense approach suggests that the student's experience with English in his native land would be a satisfactory basis for placement of foreign students in English courses at American universities. Schueler reflects the experience of many college instructors, placement advisors and administrators with the comment: "It has been found that one of the most unreliable indexes of a student's knowledge is the number of years of formal English study that he had in his native land. It has become necessary, therefore, to measure the student's command of English before admitting him to the appropriate section in the program. [The Queens College English Language Institute for Instruction of Students of Heterogeneous Linguistic Background]" (3).

The placement of foreign students in English courses became particularly acute at Columbia University which admits approximately a thousand foreign students each year for undergraduate and graduate study. The students, who come from almost every country of the world, vary in amount of education in general, in quality of preparation in the English language, and in the lengths

of time they have lived in the United States before coming to Columbia University. Because of these wide differences, it became necessary to provide English courses at levels appropriate to the mastery of English these students have.

At Columbia University, two placement tests, one written and the other oral, were developed as aids in the placement of students in the appropriate English courses. Two course sequences are provided for foreign students, one for written composition, grammar and reading, and the other for oral composition and conversation. The written test is for placement of students in the writing sequence; the oral test for the speaking sequence.

The present study reports on the relation between the written test and success of foreign students in courses in writing and reading.

THE ENGLISH EXAMINATION FOR FOREIGN STUDENTS

Two parallel forms of an *English Examination for Foreign Students* were developed by the Institute of Psychological Research at Teachers College, Columbia University. The current forms of the tests have been revised on the basis of extensive item analyses involving thousands of students: test-retest, and test-alternate test, correlations indicate that both forms are reliable.

The test, with its four multiple-choice subtests, requires sixty minutes for administration. To minimize confusion in answering, foreign students mark their answers in the test booklet by drawing a circle around the correct response.

Parts One and Two are a thirty-five item and a ten-item subtest, respectively, measuring English usage. The items are based upon actual usage errors made by foreign students. Part Three is a sixty-item vocabulary test, and Part Four is a reading comprehension test of three brief passages each of which is followed by five four-choice questions. The total score (maximum 120 points) over all four points is obtained for each student. A more detailed description of the *English Examination for Foreign Students* is given in a recent article (2).

ENGLISH COURSES FOR FOREIGN STUDENTS

The hierarchy of courses in writing and reading for foreign students in the School of General Studies of Columbia University

follows:

Name	Title	Semes- ters	Points
G.S. English eV1	Introductory English for Foreign Students	1	4
G.S. English eV3 and eV4	Intermediate Written Composition, Grammar and Reading	2	4, each semester
G.S. English V9	Advanced Written Composition, Grammar and Reading	1	4
G.S. English A3	English Composition for Foreign Students (equivalent to the freshman course in English composition)	1	4

COURSE PLACEMENT ON THE BASIS OF THE ENGLISH EXAMINATION FOR FOREIGN STUDENTS

Cut-off scores were established for the placement of students into the five English courses. The primary criterion is the score on the *English Examination for Foreign Students* (EEFS). Advisors, however, in considering other information about the students, do not adhere to the established cut-off points. Frequently, indeed, students are placed in a course despite the fact that their EEFS scores are above or below the established cut-off points for that course.

The cut-off points¹ established for eV1 Introductory was 0-40; for eV3 Intermediate 1, 41-55; for eV4 Intermediate 2, 56-70; for V9 Advanced, 71-90; and for A3 Freshman English, 91-120.

THE PRESENT STUDY

The present study is a report on the value of the *English Examination for Foreign Students* in predicting success in English courses, where success is defined in terms of the final grade assigned in the course. Final grades cover an eleven step range: A+, A, A-, B+, B, B-, C+, C, C-, D, F.

For three hundred and seventy four students, who had taken the *English Examination for Foreign Students*, final grades were available on completion of one of the five courses. Only the initial course

¹ These cut-off points have been modified since the present study was initiated.

placement was considered. The relation between letter grades, considered as eleven steps from A+ to F, and EEFS tabled for the five courses gives the essential information. In addition, the final grade distribution is reported for each course for those students whose EEFS scores were within the established range for the course and for those students whose EEFS scores were above or below the established range.

CORRELATIONS BETWEEN EEFS SCORES AND FINAL GRADES IN ENGLISH COURSES

The product-moment correlations between EEFS scores and final grades in English courses are given in Table 1 for the five courses. These range from .31 to .65.

These obtained validity coefficients resemble those of other investigators for English speaking students. Huddleston reported median correlations between four different objective and subjective test measures of English writing competence and freshman English grades as ranging from .23 to .46, with the median correlation between an objective English test and course grades as .34. In a more intensive study, Huddleston reported a correlation between an objective English test and high-school English grades of .60. In her review of the literature on objective English test prediction of final English course grades, Huddleston indicated that the range of validity coefficients was from .11 to .73, with the median about .50.

The obtained validity coefficients for the foreign students are based upon restricted ranges. The EEFS scores for students in any

TABLE 1.—ENGLISH EXAMINATION FOR FOREIGN STUDENTS: VALIDITY COEFFICIENTS BETWEEN SCORES AND FINAL GRADES: BY ENGLISH COURSE

Course and Level	No. of Students	Obtained Correlation	Correlation Corrected for Curtailment of Ranges
eV1 Introductory	63	.38	.45
eV3 Intermediate 1	76	.40	.60
eV4 Intermediate 2	34	.65	.87
V9 Advanced	110	.31	.56
A3 College freshman	91	.45	.76

course are not distributed over the entire range of EEFS scores. Such curtailment of range on one variable tends to reduce the obtained validity coefficient. To estimate the validity for the full test range the correlations were corrected for restriction of range on one variable. (4) The corrected validity coefficients are presented in Table 1.

Thus, either on the basis of the obtained, or of the corrected, validity coefficients, the EEFS seems to result in predictions as satisfactory as those for objective tests of English for native English-speaking students.

However, while the ability in English of the native English-speaking student may be considered stable by the advanced high-school or college level, that of the foreign student who has just begun his study of English is probably at its period of greatest change. The changing proficiency in English of foreign students during this period, and the vast individual differences in potential improvement (dependent upon intelligence, similarity between native language and English, experiences in the United States with English-speaking people, etc.) would tend to lower the validity coefficients between EEFS scores and final grades in English. Thus, despite such limitations, the validity coefficients between EEFS scores and final grades seem satisfactory. (1)

GRADE DISTRIBUTION IN ENGLISH COURSES FOR STUDENTS SCORING
ABOVE, WITHIN AND BELOW ESTABLISHED CUT-OFF
SCORES ON THE EEFS

The final grade distributions for students with EEFS scores within and outside the established range for each course are presented in Table 2.

In general, the final grade distribution for students within the established EEFS cut-off scores ranges from A to F: the largest proportion receive B's, and considerably fewer get A's and C's. D's and F's vary from none to nine per cent for the different courses. The exception is the first course: G.S. English eV1, in which seventy-eight per cent of the students receive A's and the others B's.

For students with EEFS scores above the established upper limit for the particular course to which they were assigned, the general tendency is for a majority of students to receive A's with relatively few receiving C's. Only G.S. English eV3, the first intermediate course, is an exception to this pattern with twenty-three per cent of

TABLE 2.—THE ENGLISH EXAMINATION FOR FOREIGN STUDENTS: THE RELATION BETWEEN ESTABLISHED CUT-OFF SCORE RANGES AND FINAL GRADE FOR SPECIFIED COURSES: FOR ALL SCORES

Course	EEFS Score Range	No. of Students	Final Grade Distribution in Per Cent			
			A	B	C	D and F
eV1	41-99	19	78	22	—	—
	0-40*	44	43	39	14	4
eV3	56-85	36	31	47	17	6
	41-55*	27	11	74	11	4
	11-40	13	—	31	46	23
eV4	71-90	2	100	—	—	—
	56-70*	16	25	63	12	—
	31-55	16	6	38	50	6
V9	91-110	12	50	42	8	—
	71-90*	64	19	45	31	5
	26-70	34	15	38	35	12
A3	91-115*	58	16	53	22	9
	46-90	33	3	36	48	12

* Established cut-off scores for the course.

the students with placement test scores above the established cut-off limit receiving grades of C or lower. This suggests that a revision of the cut-off points for this course may be needed.²

For students with EEFS scores below the established lower limit for the particular course to which they were assigned, the general tendency is for the majority of students in each course to receive grades of C or lower. There are very few A's received by this group, and the per cent of B's is usually lower than the per cent of C's.

These grade distributions indicate the value of the *English Examination for Foreign Students* in the placement of students at the appropriate level in English. The grade distribution for students with EEFS scores within the established range is satisfactory; for students with scores above the established range tends to be limited toward the upper end of the grade range; and for students with scores below the established range tends to be limited toward the lower end of the grade range.

² Such a modification has been made.

SUMMARY AND CONCLUSIONS

The validity of an *English Examination for Foreign Students* was tested against the criterion of final grades in English courses for students assigned on the basis of scores on the English test. Foreign students entering Columbia University are required to take an objective English examination on the basis of which they are placed in one of a hierarchy of five courses in written composition, grammar and reading. Final grade distributions were contrasted for students with placement test scores within and outside the established range for the course.

Obtained correlation between *English Examination for Foreign Students* scores and final grades range from .31 to .65 (corrected for range of ability, .45 to .87) validities compare very favorably with those of other investigators between scores on objective tests with final grades in English courses for native English-speaking students (1).

The final grade distribution for students with placement test scores within the established range for the course tend to be distributed over the grade range, with the largest per cent of students receiving B's and fewer receiving A's and C's. For students with placement test scores above the established limits, final grades bunched at the upper end with most students receiving A's; for those with placement test scores below the established limits, grades bunched at the lower end of the grade distribution with most students receiving C's and a considerable number receiving F's.

The data suggest that the *English Examination for Foreign Students* is a useful and valid measure for placement of foreign students in English courses appropriate to their measured mastery of English.

BIBLIOGRAPHY

1) Edith M. Huddleston. *Measurement of Writing Ability at the College-Entrance Level: Objective vs. Subjective Testing Techniques*. Princeton: N. J. Educational Testing Service, 1952.

2) Irving Lorge and Lorraine K. Diamond. "The Proficiency in the English Language of Foreign Students with Different Linguistic Backgrounds" (in press).

3) Herbert Schueler. "English for foreign students." *J. Higher Education*, 20, 309-316, 1949.

MOTIVATIONAL DIFFERENCES BETWEEN HIGH AND LOW SCHOLARSHIP COLLEGE STUDENTS

WILLIAM F. BROWN, NORMAN ABELES and IRA ISCOE

The University of Texas

A survey of the literature concerned with the study of factors influencing student success and failure in college discloses a variety of investigations (1, 3, 5, 7, 8, 9, 10, 12). Studies utilizing interviews techniques have revealed many personality differences between good and poor students. Tiebout (10), in his three-year clinical study of students at Sarah Lawrence College, found great similarities in the behavioral patterns of girls whose scholastic records were lower than intelligence test scores would predict. He concluded that his study pointed to the existence of a behavior syndrome which might be characterized as follows: (A) A need to rely upon strong and immediate motivations to begin studying; (B) A tendency to possess interests of a transitory, impermanent nature; (C) A tendency to be governed by strong hedonistic principles; and (D) A deep-seated problem in learning.

Many of the studies concerned with this problem of differences between high-scholarship and low-scholarship students have utilized the questionnaire method. Several of these reports (1, 2, 3, 5, 8, 9, 11) seem to substantiate the hypothesis that existing study habits or questionnaires do not adequately distinguish good from poor students. Brooks and Heston (1) in a study concerned with the validity of items in a study habits inventory found that the questionnaire scores correlated .18 with cumulative grade averages and .07 with ACE scores. Their sample was composed of two hundred and seventy-six college students. Of the total seventy-five items on the questionnaire, twenty-eight items were found to be significant.

Eurich (3) prepared a 100-item study habits questionnaire which he administered to one hundred and twenty-one students enrolled in a beginning course in educational psychology. The correlation between questionnaire scores and final examination grades was -.20 for the one hundred and twenty-one students. Other studies

(2, 5, 8, 9, 11) utilizing the questionnaire method for ascertaining the study habits of high- and low-scholarship students have reported similar results.

An analysis of the items on the various questionnaires seems to indicate that the items were concerned mostly with specific study techniques rather than underlying motivational differences. The studies to be presented here were designed to investigate motivational differences in terms of conformance with academic requirements and the willingness to take decisive actions. Three separate studies, each concerned with a different phase of this problem, will be discussed. We shall name them as follows: The first, or Preliminary Study; the second, or Student Attitude Questionnaire; and the third, or Classroom Study. Each of these revolves around the attendance of students at these studies. In the case of the Preliminary Study a sample of college students was asked to participate in a graduate student's research project. In the Student Attitude Questionnaire, a sample of college students was asked to attend sessions at which a questionnaire was to be administered under the auspices of Cornell University. The Classroom Study concerned a group of students taking a beginning course in educational psychology who were given the chance for additional help with their course work. This, then, briefly differentiates the three studies.

METHOD AND RESULTS

A) The Preliminary Study: Method.—The sample consisted of one hundred and forty-three second-semester freshman students, both men and women enrolled in the College of Arts and Sciences at the University of Texas. Ninety-seven students from the Dean's Honor List and forty-six students from the Scholastic Probation List were selected chiefly because they had comparable scores on the ACE (1941 Edition). The one hundred and forty-three students selected represented all the students from each list whose ACE total scores fell within certain selected limits. The upper limit (139) was based upon the highest score made by a freshman on scholastic probation and the lower limit (65) was based upon the lowest score made by a freshman on the Dean's Honor List. The difference between the means of ACE scores made by the ninety-seven students on the Dean's Honor List and the forty-six students on Scholastic Probation was not significant.

The one hundred and forty-three freshmen thus selected were

sent letters requesting that they report to one of the regular testing rooms in the Department of Psychology to coöperate in a psychological study designed to investigate sex differences in study habits. It was emphasized that the students had been randomly selected from the freshman class. They were told in the letter that the study would take about one hour and that they could come in any time during a designated three-day period to take the questionnaire. Stress was placed on the need for all subjects to appear so that the sampling design would be complete. After the initial three-day period, those who had not appeared were sent a second letter re-emphasizing the importance of their coöperation. The remaining students who had not appeared after receiving the letter for the second session were individually phoned and definite appointments were made with these students to appear at a specified time.

RESULTS: PRELIMINARY STUDY

The findings of this study are indicated in Table 1. We note that practically all of the Dean's Honor List students attended one of the sessions, while only half of the students on scholastic probation did so. Of the three students on the Dean's Honor List who did not attend, one was in the hospital and two asked if attendance was compulsory. When told that they would not be forced to appear they stated that they would be too busy to attend. Of the non-attending students on scholastic probation, only one admitted having received the two letters (i.e., when contacted over the phone) despite the fact that none of the letters were returned by the Post Office. All of the Probation List students who failed to appear had been contacted three times by phone and had made three separate appointments to appear. It is also interesting to point out that more

TABLE 1.—ATTENDANCE OF HIGH- AND LOW-SCHOLARSHIP FRESHMEN AT PRELIMINARY STUDY

Scholarship Group	Number of Students	Per Cent Attending after First Letter	Per Cent Attending after Second Letter	Per Cent Attending after Phoning	Total Per Cent Attending
Dean's List	97	80% (78)*	15% (14)	2% (2)**	97% (94)
Probation List	46	13% (6)	15% (7)	22% (10)	50% (23)

* Figures in brackets refer to actual number of students attending.

** This low percentage should be considered in the light of ninety-five per cent of the Dean's list having already taken the questionnaire.

than three-fourths of the Dean's Honor List students appeared at the first session, while only a little more than one-eighth of the scholastic probation group appeared at this session. Furthermore, the largest number of the low-scholarship students to appear did so only after phone contact.

B. The Student Attitude Questionnaire: Method.—This study took place in April, 1952, when the University of Texas was asked to coöperate with ten other universities in a study of student attitudes concerning the draft, the Korean War, college life, vocational aspirations, religious ideas, government and politics, etc. Accordingly, an alphabetical roster of all undergraduate male students at the University of Texas was prepared and every seventh student was sent a letter signed by the President of the University. This letter asked that the student attend any one of the twelve sessions scheduled during a specified two-day period. The student was informed of the nature of the questionnaire and also that he had been randomly selected. Coöperation was strongly urged as well as the fact that other universities were participating in this project so that the results would reflect upon his school. In general the letter was so worded as to make the student feel privileged to participate, rather than making this seem like a distasteful task that had to be done.

After the first session the procedure was similar to that used in the previous design. A second, more strongly worded, letter was sent which 'ordered' the student 'to report' to any one of four specified testing sessions. The students who had not appeared at the second session were individually contacted at least once over the phone and appointments were made with them to attend special testing sessions. A point was made to talk with the individual student concerned in addition to leaving messages with his roommate, landlady or housemother. This study, then, comprises a total of three hundred and sixty-seven students; namely, those who attended the first sessions (two hundred and twenty) and those who were contacted by individual phone calls, whether they eventually attended one of the sessions or not (one hundred and forty-seven). For the purpose of comparison in this study, the students who attended one of the session in response to the second letter, as well as those who in one way or another could not be contacted (i.e., moved, drafted, sick) were eliminated.

A brief listing of the categories may simplify the situation:

	Number	
Total number of letters sent to original sample*	660	
Students attending first scheduled sessions (first letter)	220	Group A
Students attending second scheduled session (second letter)	243	Group B
Students attending third scheduled session (phone contact)	81	Group C
Students contacted (phoned) but not appearing	66	Group D
Students eliminated from study (withdrew, ill, etc.)	53	Group X

* Record of attendance was available in that each questionnaire booklet contained an attendance slip. Upon completing the questionnaire a student wrote his name on the slip and turned it in separately. By noting the session attended a record was thus available while the anonymity of the students' answers was preserved.

Thus, to recapitulate, a comparison between groups A and CD (combined) was carried out in terms of grade-point averages. These were computed on the basis of $A = 4$, $B = 3$, $C = 2$, $D = 1$ and $F = 0$. This information was taken directly from the records in the Registrar's Office.

RESULTS: STUDENT ATTITUDE QUESTIONNAIRE

The findings of this study may be found in Table 2. A chi-square analysis was carried out and indicated a difference between Group A and Groups C and D which was significant at the .001 level ($X^2 = 17.38$, d.f. = 3). We may note, therefore, that the students appearing promptly after receiving their first letter had significantly higher grade averages than those who had to receive two

TABLE 2.—GRADE POINT AVERAGES OF STUDENTS APPEARING PROMPTLY VERSUS LATE APPEARING AND NON-APPEARING STUDENTS

Grade-Point Averages	Number of Students Group A	Number of Students Groups C and D	Total
3.00-3.99	60	17	77
2.00-2.99	99	65	164
1.00-1.99	49	53	102
0.00-0.99	12	12	24
Total	220	147	367

letters and at least one phone call. If a division is made at a cutting point of above and below a 2.00 point grade average (roughly a C average) the following breakdown results:

Grade-point Average	Group A	Groups C and D	Total
2.00 and above	159	82	241
Below 2.00	61	65	126
Total	220	147	367

The chi-square analysis is significant at the one per cent level of confidence ($X^2 = 9.99$, d.f. = 1). The distribution of frequencies reveals that, for the group considered, the higher scholarship students are more willing to at least comply with requests from the university authorities than low-grade-point students.

C) *The Classroom Study: Method.*—The sample of this study was composed of students in two sections of an elementary educational psychology course and consisted of eighty male and female students. They were told at the beginning of the term that one of their three examinations that semester would be taken directly from the true-false and multiple-choice questions found in their workbook. Three weeks before the examination the students were told that they could check their workbook answers with a set of answer keys constructed by the instructor. They were thus given the chance to get the correct answers to the five hundred and twenty-three questions in the workbook. The test, was to consist of one hundred of the five hundred and twenty-three questions in the workbook. The instructor stipulated that they could come any time during a two-week period to look at the keys provided they had completed their workbook exercises. The instructor frankly admitted that the students could pretend they had answered the workbook questions by haphazardly filling in pluses and minuses and numbers although he emphasized that this would not help them study for the examination. He also admitted that some students could copy the 'key' answers from another student who had done the work and looked at the keys. Furthermore, it was not compulsory to look at the instructor's key, for this was to help the student prepare for the workbook examination, and, therefore, it was optional with the student whether he wanted to appear or not. The instructor kept track of

TABLE 3.—TEST SCORES OF STUDENTS USING THE SCORING KEY AS COMPARED TO THOSE NOT USING KEY

Scores on Test	Number of Students Using Key	Number of Students Not Using Key	Total
91-100	25	6	31
81-90	11	8	19
80 and below	9	18	27
Total	45	32	77

the students using the key by having them sign their names as they looked at the key in an adjoining office.

RESULTS: CLASSROOM STUDY

The results are shown in Table 3. We note that of the seventy-seven students taking the test, forty-five made use of the key while thirty-two did not do so officially—that is they did not come and sign out for the key. The extent that those not using the key actually looked at someone else's corrected exercises is a matter for speculation.

In comparing the two groups, a chi-square analysis was made again and the difference between the two groups was found to be significant at the .01 level ($X^2 = 13.301$, d.f. = 2). The results are of course fairly obvious by inspection. There were twice as many students scoring below 80 who did not make use of the key as those who did. And of those who scored above 90 there were more than four times as many who used the key as those who did not.

DISCUSSION: THE THREE STUDIES COMBINED

What are the implications of the findings of the three studies presented? It has been known for a long time that intelligence in itself is not the only factor decisive for high or low scholarship. What, then, are some of the characteristics of the poor student? Above all, there seems to be what we shall call here 'activity delay', that is, a lack of decisiveness of action, a tendency to procrastinate, perhaps an unwillingness to conform to academic (and possibly other) requirements, routine and regulations. Whether this is merely a lack of interest, or whether there is overt interest (as is often expressed by students) which is not backed up by the activi-

ties to accompany this interest, only further research will tell. This study seems to indicate that 'activity delay' carries on beyond the classroom. Certainly most students express interest in having their attitudes 'tested'. Particularly when a whole undergraduate school is sampled it seems especially significant that the poorer students either failed to attend or did not attend until they had received two letters and one phone call. It might be argued that these students are attempting to keep up with their work and thus do not have time to attend 'attitude tests'. Yet, this is not consistent with the findings in the classroom. Even when given a chance the poorer student does not take advantage of study aids. The results of the three studies cited would emphasize that the student's attitude towards academic life may be as important (or even more so) than specific study habits, study aids, tutorial possibilities or native intelligence. Further studies are needed in order to ascertain whether these factors carry on into other areas of student life, such as participation in social and athletic activities. Related to the above is the question of what was the underlying motivation in attending college in the first place. One aspect of this process; namely, prestige in terms of membership in select groups, is now being studied by the present authors. The question can then properly be raised as to the effect of activity delay on social group activities.

SUMMARY

A series of three studies were reported concerned with the motivational differences between high-and low-scholarship students in college. The following findings were postulated:

- 1) The poor college student is characterized by 'activity delay', i.e., a lack of decisiveness of action, a tendency to procrastinate and perhaps an unwillingness to conform to academic requirements, routine and regulations.

- 2) This activity delay is not limited to the classroom only but exhibits itself in regard to activities usually regarded as outside the classroom sphere such as voluntary participation in research studies in psychology and university-wide projects such as attitude surveys.

- 3) This study pointed toward the assumption that the poor-scholarship student does not necessarily score lower on psychological tests designed to measure intelligence, but that very often

factors of interest and motivation are primary contributors towards low scholarship.

This study further raised the question whether activity delay carries over into other aspects of student life and indicated that research was in progress to study this area.

REFERENCES

- 1) F. D. Brooks and J. C. Heston. "The validity of items in a study habits inventory," *J. Educ. Psychol.*, 1945, 36, 257-270.
- 2) N. B. Cuff. "Study habits in grades four to twelve," *J. Educ. Psychol.*, 1937, 28, 295-301.
- 3) A. C. Eurich. "An analysis of self-ratings on studiousness traits," *J. Appl. Psychol.*, 1940, 24, 493-504.
- 4) P. O. Johnson. *Statistical Methods in Research*, New York: Prentice Hall, 1949.
- 5) N. M. Locke. "The student skills inventory: a study habits test," *J. Appl. Psychol.*, 1940, 24, 493-504.
- 6) Q. McNemar. *Psychological Statistics*, New York: John Wiley & Sons, 1949.
- 7) L. C. Pressey. "What are the crucial differences between good and poor students?" In S. L. Pressey (Ed.) *Research Adventures in University Teaching*, Bloomington: Public School Publishing Co., 1927, pp. 4-10.
- 8) C. W. Reeder. "Study habits," *Sch. and Soc.*, 1935, 42, 413-415.
- 9) G. M. Schlessner and C. W. Young. "Study and work habits," *Sch. Rev.*, 1945, 53, 85-89.
- 10) H. M. Tiebout. "The misnamed lazy student," *Educ. Rec.*, 1943, 24, 113-129.
- 11) A. E. Traxler. "Some results of an experimental survey of the study habits of independent-school pupils," *Educ. Rec. Bull.*, 1945, No. 42, 40-49.
- 12) C. G. Wrenn and W. J. Humber. "Study habits associated with high- and low-scholarship," *J. Educ. Psychol.*, 1941, 32, 611-616.

THE PROBLEM-ORIENTED APPROACH TO TEACHING PSYCHOLOGY

W. J. McKEACHIE and WESLEY HILER

University of Michigan

How can students comprehend and retain more of their textbook material? Many students who take a course involving concepts unfamiliar to them have considerable difficulty in gaining a concrete understanding of the material presented in their textbook. They read along passively and feel that they understand what they are reading but when examined on it, they discover how little they have mastered. We cannot attribute this difficulty merely to poor study habits, for many of the students who do poorly in a new subject are doing well in other fields with which they are more familiar.

We feel that the basic problem in these cases is the lack of an adequate frame of reference with which to organize and assimilate the various concepts, principles, and facts which are found in the textbook. Learning is easiest when it consists of filling the gaps in a "generalized knowledge of the whole."¹ Every subject matter, be it science, literature, or the arts, is an organized body of knowledge, not a mere array of isolated facts; hence a knowledge of this subject matter should be an organized structure within the student's mind. Expectations, questions, and problems are intrinsic to all such organized structures. These represent the gaps. Material which answers questions implicit or inherent in an organized knowledge of a subject matter is learned with little effort. This is in contrast to the extremely laborious and energy consuming process of rote memorizing.

On the basis of this theory, how can we help those who have no frame of reference which can be used to facilitate learning of new material, and who do not even know how to distinguish the essential points from illustrative details? If we are not to resort to rote memory and all its disadvantages, we must artificially construct a 'demand structure' in the student's mind which the facts to be learned will satisfy. This can be done by proposing a problem or

¹ This conception of learning is as old as Herbart, but has been further developed by the Gestalt psychologists.

asking a question for which the person must actively seek the answer. Preferably the problem should be meaningful to the individual and capable of arousing a genuine interest on his part. If we can induce our students to become 'problem-oriented' towards their subject matter, they will learn it more easily.

Hulsey Cason has reported the use of a work sheet system which he calls the 'intelligent question' method in teaching elementary psychology and reports great success (3). He believes that the essential thing to be learned in psychology is the psychologist's particular way of dealing with problems. Students must have an opportunity to gain practice in the use of psychological concepts and examinations must test this ability. "The students should know what kind of psychological processes will be valued and credited on the examinations, and they should be given every possible opportunity and aid in deliberately cultivating these psychological processes in their work for the course" (3, p. 363).

This procedure is recommended by Cason: "The instructor makes out the kind of questions that he desires for each topic. The answers to these questions include everything the students are asked to learn from the lectures, reading, class discussion and from any and all other sources. The one hundred or one hundred and fifty questions on the subject matter of the whole course are given to the students in advance" (3, p. 367). In this fashion the student will know what he needs to know while he is reading the text or listening to a lecture. The questions used on the examinations are derived from this list but are not identical in form.

Unfortunately Cason presents no experimental evidence in support of his method, so we are forced to generalize from the favorable results obtained with work books in high-school courses reported by Anderson (1), Harrington and Lippert (6) and Hurd (7).

An experimental study dealing with the use of work sheets is reported by Greene (4). He found on both immediate testing and testing after from one to four weeks that the results of 'guided reading' (consisting in answering true-false questions on a work sheet while reading) were very superior to either unguided reading or the lecture.

Motivational conditions in Greene's experiment, however, may have differed from those in an ordinary classroom situation. A further disadvantage is that Greene's criterion was completely contaminated, for he merely used the examination sheets for work

sheets. This procedure is somewhat similar to that of improving one's 'intelligence' by studying intelligence test manuals. Furthermore, one cannot be certain that the over-all knowledge of the guided reading group was superior to the other groups. It might be that in learning more thoroughly the material to be tested on the examination the students learned less of the other important material covered by the text but not tested.

In order to gain information concerning the value of study questions for an elementary psychology course in an actual classroom situation using items on the criterion test which are different from the study questions, the following Experiment 1 was conducted.

EXPERIMENT I

The study was carried out with three hundred and sixty students of elementary psychology at the University of Michigan. These students all used the same textbook and attended the same lectures. They had been enrolled on the basis of student choice of hour in eleven discussion sections, conducted by advanced graduate students. These discussion sections met for one two-hour period. In addition the students spent two hours a week in large lecture sections.

Seven discussion sections were given a work sheet containing seventy-five short answer questions and problems covering the textbook² material on sensation and perception paralleling the sequence of topics in the text. Four of these sections, which we shall call Group A, were not required to work out the work sheet and turn it in. The other three sections, which we call Group B, were required to work out and turn in the work sheet. The work sheets were then corrected and returned to the students. No students were discarded from Group B because of failure to do the work. As it turned out, this experimental group contains eighteen students who did not actually work out the work sheet. The remaining four sections, Group C, did not receive a work sheet.

Sections were assigned to these three groups in such a way that the three groups were as nearly equal as possible as measured by performance on the mid-term examination. Of the four instructors involved, three taught three sections each, and each of the three sections received a different treatment, while the fourth instructor

² The textbook used was Woodworth and Marquis' *Psychology*.

taught only two sections which were assigned by chance to conditions A and C.

The regular final examination (multiple-choice type) which had been prepared independently of the questions used on the work sheets served as a criterion for measuring the effectiveness of the work sheets. As it turned out only four questions on the examination covered the textbook material on perception and sensation and no examination questions were identical to those on the work sheet.

Each student in the two experimental groups was matched with a student in the control group on the basis of his score on the final examination as a whole.³ This was done by matching as many identical scores as possible and then matching scores with a difference up to 3 in such a way that the total score for each experimental group was identical to the total of the control group. Thus the means of the groups to be compared were perfectly matched. Since some cases could not be matched with any in the control group, we have an N of 87 pairs in the matching of Group B with C and an N of 111 pairs in the matching of A with C.

The number of errors each student made on the perception and sensation questions was then tabulated and the difference between pairs was obtained.

RESULTS OF EXPERIMENT 1

The mean and standard deviation of the distribution of differences of scores on the perception items was calculated, and the mean difference between the pairs was tested for significance. The difference between the group which had been given the work sheet but not required to do it and turn it in (Group A), and the control group proved to be insignificant. However the difference between the control group and the group which had been given the work sheet with instructions to work it out and turn it in (Group B) proved to be highly significant. The control group made forty-two per cent more errors on perception and sensation than did experimental Group B. See Table I.

We may tentatively conclude that under the circumstances of

³ Since only four questions out of one hundred and twenty dealt with sensation and perception, it was not necessary to subtract the score on these items from the total score in matching the groups. Our procedure would tend to diminish the significance level, but only slightly.

TABLE I.—ERRORS ON PERCEPTION ITEMS ON FINAL EXAMINATION

	Mean	Standard Deviation of Differences	N	P
Group A matched with Group C	1.063 1.252	1.16	111	.085
Group B matched with Group C	0.839 1.195	.98	87	.01

this study the use of work sheets was effective and that the group of students required to work them out was superior in the retention and comprehension of the textbook material as tested by final examination to students who had not used work sheets. Our original hypothesis about the effectiveness of work sheets has been confirmed.

STUDENT OPINION ON THE USE OF WORK SHEETS

In order to see what the students who had been given the work sheets felt about their value, a questionnaire was handed out to one of the sections which had filled out and turned in a twenty-question work sheet.

Out of the twenty students (all of the members of the class) who answered and returned the questionnaire, eighteen expressed approval of the general idea whereas two were opposed to it.

Some typical comments made by the students were as follows:

"I think they greatly aid study and help the student pick out the important points."

"These sheets definitely help in giving a clearer understanding of the textbook."

"Work sheets are an excellent idea, if they are not to be corrected, since they indicate what material is to be concentrated upon. To know what parts of the text are important (which is often difficult to determine) would bring the whole course into a clearer and better perspective."

Thus in spite of the additional work required by the use of work sheets, many students reacted favorably towards them.

EXPERIMENT 2

At the conclusion of Experiment 1, we still were left with two questions about the use of study questions.

1) Had the study questions been effective simply because they motivated the students to spend more time in studying the general area covered by the questions?

2) Is the increased learning produced by study questions gained at the expense of material not covered by the questions?

To answer these questions another experiment was carried out in which study time was controlled by giving study questions during a class period. Students spent one class period studying the topic 'Problem-solving' using thirteen study questions provided for them. A control group was instructed to spend the class period in studying the same material, but was given no study questions. The questions used on the work sheet were of the 'short answer' type, i.e., they could be answered in one or two sentences.

During the last ten minutes of the class period, the students were given a criterion test made up of ten multiple-choice items. The questions on this criterion test were constructed to fall into two categories: (1) those questions directly related to a study question (covering the same material but different in structure), and (2) those questions not based on the study questions.

RESULTS OF EXPERIMENT 2

The results of Experiment 2 showed that the students who had been given study questions did make higher scores than the control group on the items directly related to the questions. This higher score was not gained at the expense of areas not covered by the

TABLE II.—ERRORS ON CRITERION TEST

	Group Using Study Questions			Control Group			
	M	SD	N	M	SD	N	p
Questions directly related to study questions. Total possible—5.	1.9	1.2	95	2.5	1.0	85	.01
Questions not based on study questions. Total possible—5.	1.8	1.15		1.8	1.27		

questions, for scores on items not based on the study questions did not differ for the two groups. See Table II.

DISCUSSION AND CONCLUSIONS

On the basis of the evidence at hand, we conclude that work sheets are effective in improving performance on examinations. We have discussed the theoretical basis for this effectiveness in our introduction, but our experiments and the reports obtained from the students enable us to become more specific.

Work sheets help the student in the process of selection. Faced with a mass of facts, he needs to know which facts or units of material are most important. With a work sheet he can emphasize these facts in his study. Our evidence indicates that the selection of important facts does not interfere with the learning of material not selected for study questions. In addition, study questions can provide a means whereby the student can organize the vast array of facts and concepts into a unified framework which facilitates subsequent learning. The questions provide a framework into which the student can fit unfamiliar subject matter. By helping him find the logical inter-relationships between facts and principles, they may aid him in categorizing and unitizing the subject matter. Our students reported that in addition to their selective function, the work sheets give a clearer understanding of the textbook and bring the whole course into perspective.

Our theory of the function of study questions possesses certain implications for the construction of such questions.

- 1) Questions should be designed which require the student to see the implications of a fact, to relate it to other facts and to draw inferences. We would expect a question requiring only rote memory of a specific fact to be of little value in organization.

- 2) No question should be asked which requires a more complete frame of reference to answer intelligently than is possessed by the beginner in a new field. The questions should be capable of being answered correctly on the basis of the material available to the student. We would guess that a too difficult question would be likely to cause the student to distort his cognitive organization of the material in order to get some sort of answer.

- 3) No question should be asked which can be answered by a mere quotation from the book. The most effective type of ques-

tion is probably one which presents a concrete situation or problem and asks the student to interpret or solve it on the basis of certain concepts or principles. This sort of question, we believe, forces the student to look for relationships.

4) One problem which has not been touched upon is the optimal number of questions to be used for any given amount or type of subject matter. If there are too few questions, not enough of the subject matter will be covered. If there are too many questions, the factor of 'mental dazzle' (9) will enter into the picture. If there are a great number of questions and a great many new concepts, the learner will be 'dazzled' and learning will be inefficient. We would guess that the optimal number of questions would vary with the number of new concepts in the material and with the learner's familiarity with the material.

SUMMARY

1) On the basis of the theory of learning presented, a problem-oriented method was recommended for use in introductory courses.

2) Two experiments were undertaken to obtain evidence of the effectiveness of work sheets in an elementary psychology course.

3) The results proved highly significant in favor of the group using the work sheets.

4) A survey of student opinion indicated a favorable attitude towards the use of work sheet assignments.

5) Results of the experiments are interpreted in terms of value of study questions in selecting and organizing important principles.

REFERENCES

- 1) F. C. Anderson. *An Experiment To Determine the Effect of a Work Book on Achievement in General Science* (unpublished). Abstract in *Rev. Educ. Res.*, 1942, 12, 383.
- 2) H. K. Bennett. *The Iowa Plan for Directed Study through Work Type Reading*. Iowa State Department of Public Instruction, 1939.
- 3) H. Cason. "An intelligent-question method of teaching and testing." *J. Genet. Psychol.*, 1939, 54, 359-390.
- 4) E. B. Greene. "Certain aspects of lecture, reading and guided reading." *Sch. & Soc.*, 1934, 39, 619-624.
- 5) T. V. Goodrich. "Is the work book a necessity or a luxury?" *Sch. Exec.*, 1931, 50, 359-361.
- 6) E. E. Harrington and D. E. Lippert. "The work book." *Penn. Sch. J.*, 1934, 82, 359-362.

7) A. W. Hurd. "The work book as an instructional aid." *Sch. Rev.*, 1931, 39, 610-616.

8) G. Katona. *Organizing and Memorizing*. New York: Columbia Univ. Press, 1940.

9) D. Katz. *Gesalt Psychology*. New York: Ronald Press, 1950.

10) R. M. Tryon. "The development and appraisal of work books in the social sciences." *Sch. Rev.*, 1938, 46, 17-31.

HOW STUDENT-TEACHERS IDENTIFY RESPONSIBILITY IN CHILDREN

DALE B. HARRIS

Institute of Child Welfare, University of Minnesota

How do teachers define, behaviorally, responsibility and a lack of responsibility in children? To supply an answer to this question, two hundred and fifteen women education students were queried after some four months' experience in practice-teaching. The following paragraph was supplied to each student teacher:

"As a part of a research project in the development of responsibility in children, will you assist us by describing two children with whom you are now acquainted? Among the children you are now teaching, select the one you consider the most responsible and the one you consider least responsible in his or her general relationships to you, to the group or class, and to other children with whom he associates in play. For each child, write a description to show wherein he is responsible or irresponsible. Describe the behavior which has caused you to make your judgment. Just what does he do, which makes you name him as the most responsible or the least responsible child you are now dealing with? Please emphasize behavior incidents; general trait characteristics are useful but not as helpful as specific instances of behavior which you can mention."

When teachers were asked to define 'responsible' and 'least responsible' children, no requirement as to sex of child was mentioned, and, since teachers were unevenly distributed over pre-school age groups and through the elementary grades, unequal numbers at different ages were obtained. Two hundred and twelve nominations for 'responsible children' and one hundred and ninety four nominations for 'least responsible' children were collected and examined. Boys and girls appear in a ratio of about three boys to two girls. Table I shows that practice-teachers, themselves females working in the elementary grades, name boys in greater numbers than girls as being 'least responsible'. Among the younger children, girls are consistently more often selected than boys for nomination as 'responsible' children. The difference in proportions of the sexes disappears in the intermediate grades. It could be said, perhaps,

TABLE I.—NUMBERS OF CHILDREN REPORTED BY STUDENT-TEACHERS AS MOST AND LEAST RESPONSIBLE, BY AGE

Age Group	Most Responsible				Least Responsible				Total	
	Boys		Girls		Boys		Girls		N	%
	N	%	N	%	N	%	N	%		
2-3	8	20	13	32	16	39	4	10	41	100
4	3	11	11	41	10	37	3	11	27	100
5	15	24	19	31	22	35	6	10	62	100
6	14	26	17	31	16	29	8	15	55	100
7	5	27	18	40	17	38	5	27	45	100
8	7	21	10	30	9	27	7	21	33	100
9	9	26	7	20	15	43	4	11	35	100
10	13	30	10	23	16	37	4	9	43	100
11	10	36	5	18	11	39	2	7	28	100
12	7	27	6	23	12	46	1	4	26	100
13	3	27	2	18	4	36	2	18	11	100
Totals	94		118		148		46		406	

responsibility can be achieved by either sex, but girls are considerably less likely to be called irresponsible.

An attempt was made to read a number of protocols and to devise a series of categories, under which specific behaviors or behavior elements mentioned by the teachers could be subsumed. This procedure did not give 'reliable' results; i.e., too much inconsistency of classification occurred. Protocols were then read one at a time and each separate characteristic mentioned for the responsible child was marked with red pencil and listed on a large sheet of paper. Where another teacher mentioned the same characteristic in the same or almost identical words, an additional tally was made. This procedure resulted in a great number of specific items, with relatively few multiple entries. The separate items (phrased in precisely or almost precisely the same wording) were then grouped; these phrasings were collected into larger groupings by a team of judges according to agreement on manifest content. All the tallies relating to these groupings were summed and converted into a percentage of the children in that age sample. Examination of these groupings showed certain rather gross dimensions of the behavior thus classified which could be named by the judges. For the charac-

teristics attributed to most responsible children, these dimensions were: conformity, work habits, independence of action and motivation (initiative), social behavior and skill, and general personality and character qualities. A similar set of entries could describe all of the separate groupings of descriptive items for the least responsible children.¹

An example of the classification may clarify this procedure and illustrate the kinds of entries tabulated:

Conformity.—(a) Can be reasoned with; responds to suggestions; follows instructions, pays attention; listens. (b) Remembers what to do; doesn't need reminding; stays within bounds; seldom needs discipline. Can be depended on; doesn't interrupt, misbehave, run about, talk, etc; well-mannered. (c) Helps keep other children 'in line'; helps other children 'keep in line'. (d) Conforms to specific routines, involving self-help.

Corresponding entries for the least responsible group are: (a) Fails to observe rules and routines; not compliant; will not take suggestions; ignores discipline. (b) Doesn't follow directions; must be reminded; pays no attention. (c) Short attention span; distractible.

Since tabulations were originally under such specific headings, and only later collected into broader categories, it was possible to achieve considerable reliability of judgment in classification. Two judges identifying and tallying the behavior elements for a small sample of protocols achieved about ninety-five per cent agreement. Agreement in collecting these specific entries into broader categories was done by consensus of a team of three.

Not too much significance can be attached to the trends reflected by the age breakdowns in the original tables. It was thought at first that a number of such trends might appear, and indeed some few did appear which will be discussed. However, the tables were noteworthy for the fact that no characteristics appear suddenly or

¹ Detailed tables of results referred to in this paper have been deposited as Document number 3974 with the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. A copy may be secured by citing the Document number and by remitting \$1.25 for photoprints, or \$1.25 for 35 mm. microfilm. Advance payment is required. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress.

disappear completely within the age range encompassed in this study.

The responsible child is seen as a conforming child in a good many instances. Conforming to specific routines, and responding to suggestions are mentioned by teachers of pre-school groups, mainly, and reflect the importance attached in early childhood education to self-help, and socialization by learning group routines and by responding to reason and suggestion. A similar form of behavior, noted among older as well as younger children, is described by expressions like: remembers what to do, doesn't need reminding, seldom needs discipline, stays within bounds. The characteristic of helping keep other children in line, or helping children control their own behavior, is more often seen in the child between the ages of five and eight. In general, some aspect of conforming behavior, as illustrated by the phrases above, is mentioned consistently for one-fifth to one-third of the children in any age group.

Work habits are mentioned in most descriptions of responsible children. Doing good, efficient work promptly, finishing work begun, asking for extra work or doing more than is requested assume major importance early in the primary years, being mentioned for three-fifths of 6 and 7-year olds, and about ninety per cent of 8- to 11-year olds. Doing housekeeping chores and being a dependable messenger or room helper are mentioned in about one protocol in four, while putting away toys and materials appears much more frequently in the preschool and early primary years. Undoubtedly, the characteristic of being helpful to teacher is an important factor in the child's being considered responsible.

Teachers recognize independence of action as an indicator of responsibility. The child who will initiate his own activities, work independently or not need supervision, is seen as responsible in one-fourth to one-third of the cases. The qualities of being interested and highly motivated in work is more commonly mentioned after the primary years, possibly because the early enthusiasm of children for school is beginning to wane and the pupil who retains his enthusiasm is particularly appealing to the teacher. This characteristic is mentioned in about one protocol in five by ages ten and eleven.

At a time when educators are devoting a great deal of attention to social behavior and skill, it is not surprising that social-mindedness is mentioned in half or more of the protocols as an aspect of

responsible behavior. Sharing toys and taking turns is more a matter, again, of the preschool group. Contributing ideas to the group discussion, and being ready to discuss assumes increasing importance (from twenty-three per cent at four and five years to eighty-three per cent at twelve and thirteen years) in the elementary school period. Being 'very much a part of the group', fair, good sport, etc. are mentioned in one protocol in every five or six. Holding specific offices in class appears in one-third to two-thirds of the protocols.

A number of general characteristics were rather frequently mentioned for responsible children which have not been subsumed under the headings already discussed. Although maturity of behavior may be inferred from many of the earlier descriptions, the general terms mature, poised, coöperates, stable, or friendly appears in about one-fifth of the protocols. Occasionally the characteristic of responsibility is associated with being intelligent or bright. Not infrequently such terms as neat, precise, perfectionistic, or very conscientious appear. These terms suggest that some children have achieved a reputation of responsibility with their teachers through characteristics which mental hygienists sometimes consider indicative of emotional difficulty. Children may win their teachers' approval through painstaking, meticulous work at the expense of other interests and activities.

While this study did not attempt to investigate the inter-correlations of specific characteristics, it is clear that several rather different types of behavior are viewed by teachers as characterizing responsibility in children. The child who conforms to routines and teacher's expectations, who does good work and who does it willingly, going beyond the stated requirements, is seen as responsible. The child who may be personally helpful to the teacher by house-keeping chores in the room is also marked as responsible. Teachers, however, do recognize that independence of action and high level motivation may also characterize the responsible child, as may skill in social behavior which is characterized by readiness to discuss and contribute to group meetings and by election to class offices. Probably the most frequently mentioned characteristics have to do with conformity and the quality of the child's work habits; social characteristics come next in frequency of mention.

Characteristics attributed to least responsible children seem to the writer to make a clearer picture. Lack of compliance with

routines is mentioned in about half the protocols. Just as frequent is an indication of poor work habits or unwillingness to do school work. Failure to participate constructively in group activities or being an interfering factor in group activities is also associated with little responsibility (from twenty per cent to fifty per cent of the protocols).

Quite a variety of general personality and character qualities are used to describe the irresponsible child. Relatively few such general qualities were ascribed to the highly responsible child. Such terms as nervous, fidgets, highstrung, immature, unstable, cries, whines, appear in from twenty to forty per cent of the protocols. Specific interpretations of behavior as 'desires attention', 'a show-off', 'distracts others with his behavior' are given for equal proportions of the children. Being careless, forgetful, scatter-brained, unpredictable, impulsive, or losing things mentioned not at all for youngest children, appears in seventeen per cent of the protocols at ages four and five, and rises to thirty-three per cent of ten- and eleven-year olds. Being slow, tardy, or dawdling appears at all ages in between ten and twenty per cent of the protocols. Aggressive and destructive characteristics, toward children or adults, appear rather consistently in one-fifth of the cases, regardless of age, beyond the nursery group. Such characteristics as solitary, unaggressive, unsure of self, not a leader, not liked by others are mentioned in a scattering of protocols at almost every age.

While the picture of the responsible child is generally that of a well-adjusted child, the picture of the less responsible child is even more clearly that of an emotionally and psychologically unadjusted child. He is unhappy, does not get along well with his classmates, with his work, and with life in general.

Is responsibility then a matter of habits or attitudes which may be inculcated in the school program? Do teachers teach responsibility through the giving of tasks and the encouragement of work attitudes? This is possible but the present writer believes that responsibility is more likely a matter of the general level of the child's psychological adjustment which reflects in his functioning as he approaches tasks. The happy child who has some assurance of acceptance by the teacher and by others proves to be responsible. He shows enthusiasm for his work, follow-through on tasks. The unhappy, maladjusted child, perhaps because of his preoccupation with his problems, has little inclination to follow-through on

his work tasks. He is therefore identified as showing little responsibility. Whether giving such a child specific tasks and encouraging him to follow through on them will develop a sense of responsibility, can be debated. If being given the task brings with it a sense of recognition of status which the child needs, his performance, possibly, will be improved. In many cases, however, this assignment of task or being given a 'responsibility' on which he must measure up does not relate to the child's psychological disturbance. He is given one additional assignment to which he cannot marshal a concerted attack. Another in his series of failures is likely to be chalked up against him, thus further confirming the teacher's belief that the child shows very little responsibility. This is another example of the 'vicious circle' effect in the reinforcement of learning.

These descriptions pose a problem of considerable significance for mental hygiene in education. They suggest that responsibility is a correlate of good adjustment and can be achieved by whatever will improve the child's general life adjustment and that responsibility is probably not to be sought as a skill, habit or characteristic in itself.

SUMMARY

Two hundred and fifteen student-teachers supplied behavior descriptions of the most and least responsible children in their classes. While there is some tendency for teachers to nominate more girls than boys as responsible, especially in the preschool and primary age groups, there is a marked tendency for teachers to nominate boys more frequently than girls as least responsible. Responsible children are described as conforming, orderly and mature; they carry through activities and require little supervision; they are resourceful and socially oriented, participate well and help others. Least responsible children are not conforming or compliant; they need help with work or 'prodding'; they bother other children and upset orderly group activities; they appear to their teachers as tense, immature, nervous, irritable, careless, forgetful, aggressive, rebellious, or self-centered. They present a clear picture of poor personality integration and adjustment. Responsibility appears to be an expression of quality of adjustment to demands rather than a trait, habit, or skill in itself.

FALSIFICATION OF INTEREST PATTERNS ON THE KUDER PREFERENCE RECORD

EDWARD J. DURNALL, JR.

Nasson College, Springvale, Maine

Various studies have been conducted to determine whether it is possible to falsify interests on the Kuder Preference Record. The general conclusion appears to be that it is quite simple to achieve a high or a low score in a particular interest area (1). This fact has tended to limit the use of the Kuder Preference Record in those situations where a certain interest was required for employment or for admission to an institution or a particular program of instruction. The possibility of falsifying an entire interest pattern, rather than one or two interest areas, has not been adequately explored.

The method of this study was the administration of the Kuder Preference Record-Form BB to a group of students, using the instructions furnished by the author. The inventory was then repeated within one week, with specific instructions to the students to check the items as they believed a person with interest in a certain occupational field would check them. The occupational field selected was that of accountant-auditor, since Kuder has furnished norms for this group (2) and since the accountant-auditor profile score may be determined by means of Fisher's discriminant function (3). The Otis Quick-Scoring Mental Ability Test-Form C was administered to the group to give some measure of intelligence in relation to the ability to falsify.

The group selected for this experiment was made up of two classes in personnel management at a four-year co-educational, liberal arts college. As the number of women involved was quite small, the analysis was limited to the scores of the seventy-nine men.

RESULTS

1) There is little doubt that it is possible to simulate entire interest patterns of certain occupational fields. Figure 1 presents a graphic comparison of the mean interest profiles of the students on the first testing as compared with those of personnel managers as reported by Kuder (2), and Figure 2 presents a graphic comparison

PERCENTILE SCORE

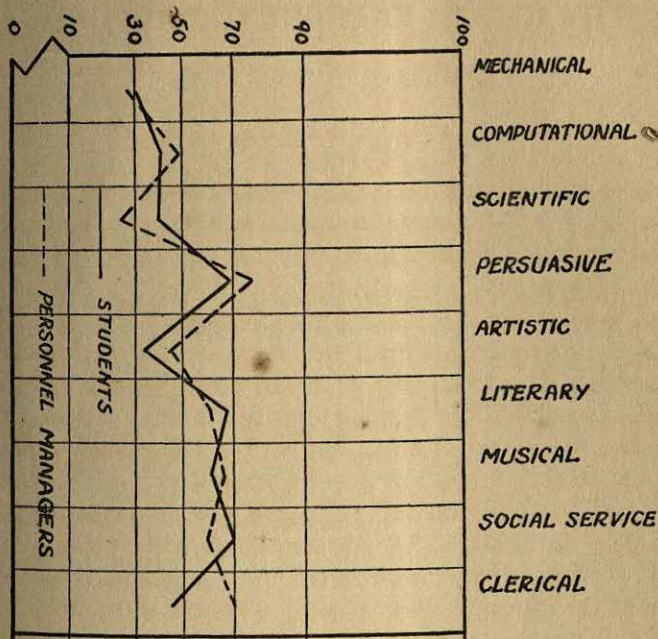


FIG. 1.—Profile of means of students on their first test compared with the Kuder Norms for Personnel Managers.

of the mean interest profiles of the students on the second testing as compared with accountant-auditors as reported by Kuder (2). It will be noted that the average true-interest profile of the group was similar to that of personnel managers while, with the simulated scores, the average interest profile of the students was typical of accountant-auditors.

2) The mean accountant-auditor score of the group on the first test was 33.25. On the second or simulated test, the mean was 103.90. The critical ratio between these two scores was 19.14, leaving little possible doubt that the difference was significant.

3) The students with high accountant-auditor scores on the first administration of the Kuder made slightly higher accountant-auditor scores on the second administration than did the students with low accountant-auditor scores. The mean accountant-auditor score of the upper quarter from the first administration was 105.63

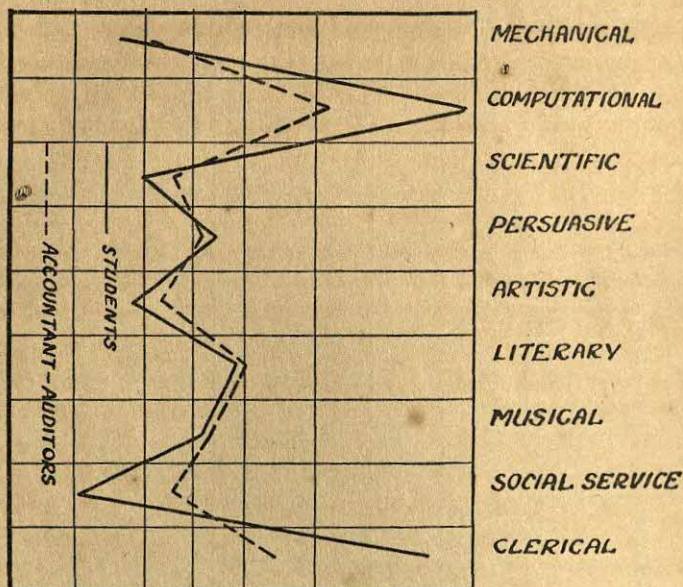


FIG. 2.—Profile of mean of students on their simulated test compared with the Kuder Norms for Accountant Auditors.

on the simulated testing; while that of the lower quarter from the first testing was 98.84 on the second. The critical ratio between these two scores was 1.09.

4) The correlation between the gain in the accountant-auditor score from the first administration to the second administration and the Otis IQ was .39, with a standard error of .09. A similar correlation (.31) was found to exist between the accountant-auditor scores on the simulated test and the Otis IQ scores. The multiple correlation coefficient between Otis IQ scores and true accountant-auditor scores on the one hand and simulated accountant-auditor scores on the other, was .50, with a standard error of .08.

CONCLUSIONS

It seems apparent from this study that it is entirely possible to falsify an entire interest pattern on the Kuder Preference Record. That the ability to do this is related to IQ is shown by scores on the Otis test. It is also related to the true interest patterns of the students, since the student possessing a high Otis IQ and some

interest in the field to be simulated finds it quite simple to achieve high scores on the profile for the occupational field in question.

The use of the Kuder Preference Record in situations which promise to reward a person for showing a certain interest pattern is open to serious question.

REFERENCES

- 1) Orrin H. Cross. "A study of faking on the Kuder preference record." *Educational and Psychological Measurement*, 10, 271-277, 1950.
- 2) G. Frederick Kuder. *Revised Manual for the Kuder Preference Record*. Chicago: Science Research Associates, 1946.
- 3) R. M. W. Travers. "The use of a discriminant function in the treatment of psychological group differences." *Psychometrika*, 4, 25-32, 1939.

THE RELATIONSHIP BETWEEN CLASS ABSENCES AND COLLEGE GRADES

ALEXIS M. ANIKEEFF

Department of Psychology

Oklahoma A. and M. College

For the purpose of developing a sound administrative policy toward college class attendance, an inquiry was initiated to determine the nature of relationships which may exist between class absences and final grades of students enrolled in the School of Business and Industry.

PROCEDURE

At the conclusion of the fall semester, thirty-four faculty members in the School of Business and Industry submitted a roster of 2654 students who were enrolled in their classes. This roster indicated the number of class absences for each student, and the student's final letter grade in the course. Established numerical quality point equivalents of 3, 2, 1, 0, and -1 were substituted for letter grades A, B, C, D, and F, respectively. Students were then placed in categories according to the number of absences which each student accumulated during the semester in each of his courses. Arithmetic means of quality point distributions were next derived for each absence category. In this manner, data became available for plotting class absences on the Y axis and quality point averages on the X axis of scatter diagrams. Linearity of scatter diagrams suggested the computation of Pearsonian product-moment coefficients of correlation.

Moreover, in order to visualize the obtained data from another perspective, students were grouped according to letter grades and the median number of absences for each letter-grade category was computed. This procedure represented more realistically the usual academic routine of assigning grades and recording number of class absences. In addition, the use of medians rather than arithmetic means, reduced the potency of extreme cases. As such, the data conceivably appeared in a more descriptive form.

Inasmuch as the same instructor indicated both the letter grade

and the number of absences for each student, an instructor's proclivity to assign grades which are based in part upon the number of absences accumulated by a student could lead to spuriously high correlations between grades and absences. As a partial measure of the extensiveness of such a practice, each instructor was ranked according to the median number of absences found in his classes. The instructor was again ranked in terms of the arithmetic mean of the quality points which he assigned as final grades to students. Spearman rho as illustrated by Guilford (2) was derived to describe the relationship between instructors ranked according to medians of class absences and instructors ranked according to arithmetic means of assigned quality points. Although the obtained correlation may not indicate the specific extent to which each instructor used number of absences as a criterion for determining final class grades, some measure of the over-all tendency for all instructors to do so could be indicated.

RESULTS

Quality points were found to vary inversely with number of absences on all class levels, as Table 1 indicates. The obtained coefficients of correlation range from $-.633$ to $-.962$, and in all cases are significant at either the .05 or .01 levels of confidence. By squaring the r obtained for the four-year breakdown, it is evident that 83 per cent of the class grade variance in this category can be attributed to number of class absences and other related factors, while only 17 per cent of the class grade variance can be attributed to unrelated factors. Using the same procedure, the reader can determine the extent of variance attributed to class absences for the remaining class levels.

Since the number of cases in each absence category decreases as the number of class absences increases, a cut-off point had to be established which would permit the greatest number of points to be plotted on a scatter diagram without sacrificing the stability of each retained point. It was found advisable to exclude from analysis all absence categories which contained less than ten cases. Table 1 indicates the percentage of population used to establish scatter diagram points on each class level after excluding from analysis all absence categories which contained less than ten cases. In this connection, it should be noted that although an absence category may be excluded in one class level breakdown, the previously ex-

TABLE 1.—RELATIONSHIP BETWEEN CLASS ABSENCES AND QUALITY POINTS

Class Levels	Per Cent of Population	N	r	$\bar{r}$
Four Year	99.13	16	-.910**	-.816
Junior-Senior	97.38	12	-.962**	-.918
Freshman-Sophomore	98.89	15	-.936**	-.867
Senior	96.54	9	-.706*	-.426
Junior	95.13	12	-.902**	-.797
Sophomore	97.31	13	-.633*	-.346
Freshman	94.86	11	-.929**	-.848

* -.01 level of confidence.

** -.05 level of confidence.

cluded absence category can be included in the combined class level breakdowns owing to the function of summation. As a consequence, the r obtained for the four-year breakdown includes absence categories ranging from no absences to sixteen absences, and excludes less than one per cent of the cases. In Table 1 the double asterisk indicates a .01 level of confidence, while the single asterisk indicates the .05 level. In the same table, obtained r 's were corrected for a small number of cases as suggested by Croxton and Cowden (1).

In general, for all cases and on all class levels, the median absence is smallest for A grade students and increases progressively for B, C, D, and F grade students. Table 2 illustrates the foregoing statement. It is noteworthy that the median absence of senior and junior level A grade students differs markedly from B

TABLE 2.—MEDIAN NUMBER OF CLASS ABSENCES ACCORDING TO LETTER GRADES

Class Levels	Letter Grades				
	A	B	C	D	F
Four Year	1.82	2.50	3.26	3.85	5.19
Junior-Senior	1.74	3.03	3.51	4.50	5.75
Freshman-Sophomore	1.90	2.02	3.07	3.63	5.04
Senior	1.35	3.00	2.95	5.67	2.00
Junior	2.19	3.06	3.45	4.25	6.00
Sophomore	2.11	2.24	3.38	4.05	5.40
Freshman	1.62	1.63	2.36	3.06	4.78

grade students on the same levels. A similar situation does not appear to hold for freshman and sophomore levels. The apparent inconsistency of the senior F grade student's absence median in Table 2, results from the use of a very small number of unevenly distributed cases.

No significant relationship is found between number of class absences and grading leniency of an instructor. Coefficients of correlation obtained for this comparison range from $-.118$ on the four-year level, through $-.139$ for the junior-senior level, to $-.367$ for the freshman-sophomore level.

DISCUSSION

A statistically significant inverse relationship is found between number of absences and final grades, for all class levels and all combinations of class levels. Although the relationship between faculty grading leniency and number of class absences is both very low and statistically non-significant, the extent to which each faculty member uses number of class absences as a partial criterion for assigning final grades remains unknown. Theoretically, if all faculty members weighted number of class absences as a criterion for assigning grades to the same degree, and if all faculty members used identical systems of grade distributions, strong negative relationships should have been obtained. In brief, instructors who had the greatest number of absences would have issued the lowest grades to their students. Actually, although the obtained coefficients of correlation are negative, these correlation coefficients are neither very strong nor statistically significant.

Another factor which could contribute to the inverse relationship between quality points and class absences, is that of motivation. Lack of motivation could explain the accumulation of a large number of class absences. Moreover, if motivation is wanting, a student expends less effort on his studies and thereby earns lower grades in his course work. As a consequence, whether one measures the number of accumulated class absences or the excellence of a student's grades, the same factor—motivation—may conceivably be measured in both cases. Two measurements of the same factor should correlate highly with each other since both are obviously measures of the same characteristic being compared with itself.

An approximation of the extent to which motivation may contribute to class grades can be secured by relating the factor of motiva-

tion to those factors which well-documented research indicates are related to college grades. Possibly the most documented factor is that of intelligence. Since numerous summaries of the relationships between intelligence tests and grades (3) for college students range from .20 to .79, with the modal coefficients ranging from .30 to .50, cognizance must be made of the intelligence factor and its contribution to the grade received by a student. This being the case, the factor of motivation can only contribute that portion to the class grade which has not already been contributed by intelligence and other factors. In order to support the assumption that the factor of motivation can account for the high coefficients of correlation which are obtained between number of class absences and honor point averages, the results obtained in this study must be reconciled with results obtained and documented in other studies. Assuming an absence of overlapping in correlations between intelligence and motivation, as well as an absence of correlation between class absences and intelligence, results obtained in this study would permit a range of correlation coefficients from .27 to .77 to exist between intelligence and class grades. Such a range of correlation coefficients approaches rather closely the previously indicated range of summaries of relationships between intelligence and college grades. Consequently, the assumption that motivation of a student can account for the high correlations found between number of class absences and honor point averages is not inimical to results obtained in other studies.

The obtained results are obviously peculiar to the sample used and the methods employed in this study. However, to the extent that other studies, using other samples, and employing diverse experimental designs, obtain comparable results, an interesting possibility presents itself. Instead of spending countless hours in the construction, development, administration, reading, and grading of college examinations, an instructor could assign grades according to number of class absences and apparently achieve closely approximate results in most instances.

SUMMARY

Thirty-four college faculty members submitted a roster of 2654 students who were enrolled in their classes. This roster indicated both the number of class absences and the final course grade for each student.

1) After letter grades were transformed into quality point equivalents, Pearsonian coefficients of correlation were derived by placing number of absences on the Y axis and quality point averages on the X axis. The obtained coefficients ranged from $-.633$ to $-.962$, and in all cases were significant at either the .05 or .01 levels of confidence.

2) The median number of absences for each letter grade category was derived for all class levels. In general, for all cases and on all class levels, the median number of absences was smallest for A grade students and it increased progressively through letter grades B, C, D, and F.

3) As a partial measure of the prevalency with which instructors may use the number of class absences for the determination of class grades, a Spearman rank order coefficient of correlation was plotted by ranking instructors according to the median number of absences in their classes and the arithmetic mean of the final class grades issued. The coefficients of correlation obtained for this relationship ranged from $-.118$ to $-.367$, and none were significantly reliable at traditional levels of confidence.

4) Since the number of accumulated class absences and the earned quality point averages may both have a motivational factor in common, the obtained coefficients of correlation between number of class absences and quality point averages may be spuriously high.

REFERENCES

- 1) F. E. Croxton and D. J. Cowden. *Applied General Statistics*. New York: Prentice-Hall, 1939.
- 2) J. P. Guilford. *Fundamental Statistics in Psychology and Education*. (2nd Ed.) New York: McGraw-Hill, 1950.
- 3) D. E. Super. *Appraising Vocational Fitness*. New York: Harper, 1949.

BOOK REVIEWS

EVERETT L. SHOSTROM AND LAWRENCE M. BRAMMER. *The Dynamics of the Counseling Process*. New York: McGraw-Hill Book Company, Inc., 1952, pp. 213. \$3.50.

This volume is a practical rather than theoretical discussion of the counseling process, with particular emphasis on the treatment of educational-vocational problems. The authors herein define counseling as "a purposeful, reciprocal relationship between two people in which one, a trained person, helps the other to change himself or the environment." The stated frame of reference is 'self-adjustive' with the aim of counseling being greater self-direction and self-responsibility. This actually represents the philosophy and techniques of Rogerian client-centered therapy with some minor modifications. The modifications, for the most part, involve recommending greater use of psychological tests and inventories than is usually advocated by Rogers and his students.

The goal of this 'permissive or client-centered approach' is "to facilitate the individual's total maturity as well as focus on his immediate problem-solving processes." Counseling is seen as a situation where the client "may reevaluate his experiences and so bring the self into closer harmony with experience." The self, in this case, is defined as "the individual's dynamic organization of concepts, values, goals and ideas which determine the ways in which he should behave."

After discussing the self-adjustive approach in terms of Roger's notions of personality and growth, the authors turn their attention to actual counseling procedures. They discuss establishing readiness for counseling, the initial interview, the use of tests and occupational information, and the synthesis (or terminal) interview. The final chapter is a discussion of the application of the client-centered approach to classroom teaching, the organization of courses and their content, and concludes with a client-centered philosophy of education.

Shostrom and Brammer, throughout the text, stress the important problem of dealing with the client's attitudinal and emotional responses to counseling. Their use of pre-counseling group orientation sessions to explain the goals of counseling and to help clients clarify their expectations about counseling seems to be a very use-

ful procedure. The section on interpreting test results to the client is very well done and may provide counselors with several helpful techniques.

The authors strongly advocate the self-adjustive, i.e., client-centered, approach as preferable to the directive approach, the only other alternative discussed. Their reasons for this position are two-fold—ideological and empirical. They believe that “the basic assumptions of democracy and those of client-centered therapy are one and the same.” Implicit in this statement is the argument that, if democracy is the best form of government (and who would argue against this point of view), then client-centered treatment is the best form of therapy. Meehl and McClosky in “Ethical and political aspects of applied psychology” (*J. abnorm. soc. Psychol.*, 1947, pp. 91–98) have pointed out the sophistry of this reasoning and called for the use of empirically derived criteria to evaluate the efficacy of the counseling process.

The advocacy of this position by the authors also stems from their participation in the Stanford Guidance Study, a research program which is reported in some detail. This study seemingly used three criterion measures of the effectiveness of counseling: (1) factual information about counseling procedures (the client's ability to differentiate different types of tests; ability to name commonly used tests; etc.), (2) feeling tone (how the client ‘feels’ about the counseling experience; to be inferred from his verbal statements.), (3) self-direction (the client's attitudes about himself in relation to a counseling experience; to be inferred from verbal statements). The authors claim that the results of the study demonstrated that the group treated by client-centered techniques was ‘superior’ on the last two criteria to a group counseled by ‘more traditional directive techniques’.

One may question the usefulness of these particular criterion measures and the authors' conclusions, based on this single study, about the over-whelming advantages of the client-centered approach. Other studies, using different criterion measures, have come to quite different conclusions. For example, in a recently completed study, Klingelhofer (*The Effects of Educational Counseling on the Scholastic Achievement of Probationary College Students*. Unpublished doctor's dissertation, State University of Iowa, 1953.) found that a highly directive approach in counseling was efficacious, using improvement in the scholastic performance of college stu-

dents on probation as the criterion measure. A survey of the research literature in this area (unfortunately neglected by the authors) would scarcely permit unequivocal statements about the over-all superiority of one counseling technique over another at this time.

This reviewer wonders about the audience for which this text was written. If directed to the professional psychologist, then much of this material such as the discussion of the learning theories of Hull, Tolman, et al, will be quite low level. If, on the other hand, it is aimed at the guidance officer(s) in the secondary school or college, then these discussions require more psychological training than such readers are likely to have. The problem of the audience seems especially acute as the authors do not discuss the problems of counselor training or counselor qualifications. This is especially unfortunate in these times of a growing professional awareness on the part of counselors.

While they do point out that these counseling procedures are usually for the individual with 'personal problems that we all have' and that students with problems 'too involved' should be turned over to a psychologist, no operational definitions of such cases are supplied nor are the possible dangers involved in such cases discussed. No differentiation is made between the counseling procedures recommended by the authors and those to be followed by the psychologist to whom the client is referred. Indeed, no mention is made anywhere in this volume of psychotherapy or of any of the other usual techniques that might be utilized by a clinical psychologist after referral.

The volume is quite readable, very practical, and filled with many 'helpful hints' about counseling and interviewing. Therein lies the danger. A naïve, unsophisticated reader may imagine that all counseling is as simple as the authors' specimen cases and can be learned, 'cook book' fashion, with the present volume as the guide. It is extremely unfortunate that the limitations of these client-centered techniques and the necessary cautions that should be followed in all counseling programs are not more fully treated.

LEONARD D. GOODSTEIN

State University of Iowa

NORMAN R. F. MAIER. *Principles of Human Relations*. New York: John Wiley & Sons, Inc., 1952, pp. 456.

Human relations problems have received considerable attention from industrialists and psychologists. Maier's book presents principles and procedures that make important contributions to solutions of these problems. The book is designed to be used by psychologists and by industrial personnel who recognize the problems to be solved but have encountered resistance to the development of better human relations. It may not have as much appeal to industrialists who do not recognize existing problems or to those who have not had enough psychological training to recognize the value of the principles and methods presented. Certainly, trainers looking for a 'recipe book' will be disappointed. However, Maier's book does offer more than enough material to develop an excellent training program if the builder is prepared to exploit what is offered and to tailor the program to the needs.

The design of the book may have placed some limitations upon its use. Maier says: "Basically, supervisory training is a preventative program in the control of attitudes." The discussion is confined to principles and methods for the changing of attitudes. The training of supervisors, on the job, is mainly a problem of this kind. However, the training of employees to take new supervisory jobs may require the development of new concepts and new behaviors as well as the changing of attitudes. Maier recognizes that new behaviors have to be developed from time to time, but he mentions that the usual training methods are adequate for this and goes no further.

Consideration of the content leaves little doubt that the book can be an important aid. Although the chapters are not grouped into sections, there is significant order given to their arrangement. The Introduction clearly sets up a philosophy of training for better human relations in industry. In Chapter Two there is an outline for a training program. The next chapters present methods to be used in training programs and group activities. Discussion methods, rôle-playing, and a group decision procedure are exceptionally well handled. Chapters Eight and Nine contain case studies which are illustrations of the content found in actual situations where the above mentioned methods were used. The last two chapters depart

from group participation and discuss, respectively, the supervisor's dealing with individuals and non-directive counseling.

The author has earned considerable respect for his work and writing. The present book will not lessen this respect. He has written for all persons interested in human relations problems and all of these people should have his book available for reference. The book is valuable because it offers so much that has been subjected to empirical trial and that is so clearly presented. The best human relations programs will always involve much of what has been dealt with in this book.

DUANE C. SHAW

Carnegie Institute of Technology

PUBLICATIONS RECEIVED

- RUSSELL L. ACKOFF. *The Design of Social Research*. Chicago: University of Chicago Press, 1953, pp. 420. \$7.50
- GORDON W. ALLPORT. *The Nature of Prejudice*. Cambridge, Mass.: Addison-Wesley Pub. Co., Inc., 1954, pp. 537. \$5.50
- KENNETH L. BEAN. *Construction of Educational and Personnel Tests*. New York: McGraw-Hill Book Co., 1953, pp. 231. \$4.50.
- BRUNO BETTELHEIM. *Symbolic Wounds: Puberty Rites and the Envious Male*. Glencoe, Illinois: The Free Press, 1954, pp. 286. \$4.75.
- GLENN MYERS BLAIR, R. STEWART JONES, AND RAY H. SIMPSON. *Educational Psychology*. New York: The Macmillan Co., 1954, pp. 601. \$4.75.
- DOROTHY R. BLITSTEN. *The Social Theories of Harry Stack Sullivan*. New York (313 W. 35th St.): The William-Frederick Press, 1953, pp. 186. \$3.50.
- TRIGANT BURROW. *Science and Man's Behavior: The Contribution of Phylobiology*. New York: Philosophical Library, 1953, pp. 564. \$6.00.
- JOHN B. CARROLL. *The Study of Language*. Cambridge, Mass.: Harvard University Press, 1953, pp. 289. \$4.75
- STUART CHASE. *Power of Words*. New York: Harcourt, Brace & Co., 1954, pp. 308. \$3.95.
- LUELLA COLE. *Psychology Of Adolescence*. Fourth Edition. New York: Rinehart & Co., Inc., 1954, pp. 712. \$6.00.
- LEE J. CRONBACH. *Educational Psychology*. New York: Harcourt, Brace & Co., 1954, pp. 628. \$5.50.
- LESTER D. CROW AND ALICE CROW. *Child Psychology: Summary of Growth and Development with Review Questions and Answers*. New York: Barnes & Noble, Inc., 1953, pp. 267. (paper) \$1.50.
- ROBERT C. CRAIG. *The Transfer Value of Guided Learning*. New York: Bureau of Publications, Teachers College, Columbia University, 1953, pp. 85. \$2.75.
- HORACE B. ENGLISH. *The Historical Roots of Learning Theory*. Garden City, N. Y.: Doubleday & Co., Inc., 1954, pp. 21. (paper) \$0.65.
- H. J. EYSEBCK. *The Structure of Human Personality*. New York. John Wiley & Sons, Inc., 1953, pp. 348. \$5.75.

- ILSE FOREST. *Child Development*. New York: McGraw-Hill Book Co., 1954, pp. 291. \$4.00.
- LAWRENCE K. FRANK. *Feelings and Emotions*. Garden City, N. Y.: Doubleday & Co., Inc., 1954, pp. 38. (paper) \$0.85.
- D. C. FRASER. *A Psychological Glossary*. Cambridge, England: W. Heffer & Sons, Ltd., 1954, pp. 40. (paper) 3s 6d.
- TIMOTHY GANNON. *Psychology: The Unity of Human Behavior*. Boston: Ginn & Co., 1954, pp. 482. \$4.75.
- GEORGE E. GARDNER. *Case Studies in Childhood Emotional Disabilities*. Volume I. New York: American Orthopsychiatric Association, Inc., (1790 Broadway), pp. 368. \$5.00.
- SISTER M. MYNETTE GROSS. *Teaching Success of Catholic Elementary School Teachers*. Washington, D. C.: The Catholic University of America Press, 1953, pp. 129. (paper) \$2.25.
- GEORGE W. F. HALLGARTEN. *Why Dictators? The Causes and Forms of Tyrannical Rule since 600 B. C.* New York: The Macmillan Co., 1954, pp. 279. \$5.50.
- FLORENCE HALPERN. *A Clinical Approach to Children's Rorschachs*. New York: Grune & Stratton, Inc., 1953, pp. 270. \$6.00.
- STARKE R. HATHAWAY AND ELIO D. MONACHESI. *Analyzing and Predicting Juvenile Delinquency with the MMPI*. Minneapolis: University of Minnesota Press, 1953, pp. 153. \$3.50.
- Health and Human Relations*. Report of a Conference on Health and Human Relations Held at Hiddesen near Detmold, Germany, August 2-7, 1951, Sponsored by The Josiah Macy, Jr. Foundation. New York: The Blakiston Co., Inc., 1953, pp. 192.
- ARCH O. HECK. *The Education of Exceptional Children*. Second Edition. New York: McGraw-Hill Book Co., 1953, pp. 513. \$6.00.
- ELIZABETH B. HURLOCK. *Developmental Psychology*. New York: McGraw-Hill Book Co., 1953, pp. 556. \$6.00.
- CHRISTINE P. INGRAM. *Education of the Slow-learning Child*. New York: The Ronald Press Co., 1953, pp. 359, \$5.00.
- PEYTON JACOBS. Joe S. Jacobs, Editor. *The Behavior Cycle: An Interpretation of Behavior from the Standpoint of an Educationist*. Ann Arbor, Mich.: Edwards Brothers, Inc., 1954, pp. 126.
- ARTHUR T. JERSILD, KENNETH HELFANT AND ASSOCIATES. *Education for Self-understanding*. New York: Bureau of Publications, Teachers College, Columbia University, 1953, pp. 54. (paper) \$0.85.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 45

May, 1954

Number 5

A GROUP-DISCUSSION METHOD OF TEACHING PSYCHOLOGY¹

ALBERT EGLASH

Mayor's Rehabilitation Committee, Detroit, Mich.

Lippitt and White's (3) experiment on democratic and autocratic leadership of boys' clubs is the indirect source of the teaching technique used in this investigation, and is the direct source of the experimental design. The direct source of the teaching method is a seminar on supervision in industry, given by N. R. F. Maier at the University of Michigan, in which the principles applied by Lippitt to boys' clubs are applied to industry. Maier defines democratic leadership in terms of the leader offering the group problems rather than solutions. Within the limits of his area of freedom, the democratic leader allows the group to make its own decisions.

Both directly through his own writings and indirectly through this seminar, Carl Rogers' (4) nondirective philosophy has played an influential rôle here. In addition, this approach is related ideologically to Anderson's (1) studies of integrative and dominative behavior in classrooms.

The experiment¹ has been designed to answer these questions:

- 1) How will achievement in a group-discussion class compare with that in a lecture class?
- 2) How will students' reactions in a group-discussion class compare with those in a lecture class?

PROCEDURE

Both classes were in Introductory Psychology, and were taught by the same instructor, who has had the same amount of experience with discussion and lecture methods of teaching. In both classes the term's reading assignment was given on the first day of class.

¹ Conducted at Michigan State College.

At the end of each two-week period, the same multiple-choice quiz was given to each class. These quizzes included both text and class material. In addition, each class was given the same comprehensive departmental final. On the last day of the term, each section was asked to complete a Teacher Evaluation Sheet used by the college.

An attempt was made in both sections to keep the emotional atmosphere warm and permissive. The instructor allowed himself to be the same person in each section—informal, considerate of students' needs, undogmatic, i.e., quick to admit his own lack of knowledge or conviction, and with humor directed more towards himself and his profession (teacher, psychologist, scientist) than towards the students. He felt, in both sections, that he had gotten to know some of the students as individual personalities, and that a very human relationship had been established between himself and the class. Students in both sections were given an opportunity to criticize the conduct of the course and to offer suggestions for improvement. In both classes some of the suggestions were utilized.

In short, the instructor attempted to keep the classes identical except with respect to teaching method. However, the frequency of meeting differed: The discussion class met for one hour a day in the afternoon, four days a week; the lecture class met for two hours an evening, two evenings a week. While one might expect the evening class to have more adults, casual observation suggested that both classes were composed almost exclusively of students of the same age group. Mean ACE scores and grade-point averages were almost identical in the two sections (respective CR's are .38 and .64). Each section held approximately fifty students.

DISCUSSION CLASS

On the first day of the quarter, the students were told that this would be a discussion class, and that no lecture material would be given. They were asked to form discussion groups of approximately six students each, and to select a discussion leader. Each group then proceeded to discuss a series of questions given them by the instructor, questions relating to the conduct of the course: number, kind, and weighting of examinations; extent to which class attendance and participation should influence the grade; topics they wished to take up. After a twenty-minute group-discussion period, each group leader reported his group's decision to the class, and the

instructor's decision, based on these eight group decisions, was then publicly made.

This same method was used in order to cover the content of the course. At the beginning of a two-week period, each group was given a typewritten question, and the groups were given time in class to discuss their respective questions. After a twenty-minute group-discussion period, each leader reported to the class the ideas developed within his group. The class as a whole would then discuss the group's ideas.

The instructor's rôle was one of listening, trying to understand each point as it was made, occasionally reflecting or summarizing what was said, sometimes integrating one comment with another, and listing the various points on the blackboard. Incorrect information was uncritically accepted and listed. Requests for facts or for opinions were handled by the instructor offering the question back to the class. While this technique resembles one which Faw (2) has reported, in that the discussion is student-centered, it differs in two respects: 1) The discussion is oriented primarily around questions devised by the instructor and 2) Intellectual content rather than emotion is emphasized.

After each of the five class quizzes, the distribution of scores was put on the board, and each discussion group recommended a cut-off point for each letter grade. After the class decision had been made, each individual was notified of his own score. After each quiz, the discussion groups were given the task of offering suggestions for improving the conduct of the course.

LECTURE CLASS

The decisions concerning examinations which had been made by the discussion section were imposed on the lecture class. The material growing out of class discussions was offered to the lecture class by means of lecture, movies, and other demonstration materials. The same number of A's, B's, etc., which the discussion class gave themselves after each quiz were given to the lecture class.

RESULTS

Achievement.—There are no consistent differences either in means or in variances, and what differences there were were not statistically reliable. On the first quiz, the examination questions and the discussion questions were on the text material. The discussion

group did somewhat better. During the second two-week period, the discussion questions related to the groups' topics. For example, one group had expressed an interest in doodling. Since the text assignment was on sensation, they were given the question, "What senses are involved in doodling and finger-painting?" Here the trend is sharply reversed. As in the other quizzes, we find little difference in achievement on material which, growing out of class discussion, was presented by lecture to the control group.

During the third period, the discussion questions were on class procedure, e.g., "What are the things you like and dislike about this method of teaching elementary psychology?" Achievement on text material, which was not mentioned in the discussion class but which was explained to the lecture class, is equal for the two groups. During the fourth period, as during the first, the discussion questions related to text material, though less directly than before. As on the first quiz, the discussion class did better on text material, but did no better on discussion material. During the fifth two-week period, the discussion questions again related to the text material. Here the two classes are equal on both text and class material.

The final examination failed to indicate any difference in achievement between the two classes.

Attitudes.—(a) Quantitative: As shown in Table 1, the five-point

TABLE 1.—TEACHER EVALUATION SHEET

Scale	Mean scores		CR	F Variance ratio: Discus- sion/Lec- ture
	Discus- sion	Lecture		
1. Were important objectives met?	3.3	4.2	5.29**	2.64**
2. Did the instructor's presentation enhance learning?	2.4	3.6	5.45**	1.86*
3. Is the instructor's speech effective?	3.5	3.8	1.76	1.12
4. How well does he work with students?	3.6	3.9	1.20	1.91*
5. Does he stimulate independent thinking?	4.1	3.7	2.00*	1.12
6. Do his grading procedures give valid results?	3.8	4.2	1.82	1.47
7. How does he rank with other instructors?	3.1	3.9	4.00**	1.42

* .05 level of confidence.

** .01 level of confidence.

scales of the Teacher Evaluation Sheet indicate differences both in means and in variances between the two groups. The lecture class felt, more than did the discussion class, that the important objectives of the course had been met; that the instructor's method of presenting the material enhanced the learning experience; that the instructor works well with students, i.e., they feel free to come to him for extra help; that the grading procedure is satisfactory; and that he should be ranked above the average instructor in the college. Only with respect to his ability to stimulate independent thinking did the discussion group rate the instructor higher than did the lecture class. Variances for the discussion class are consistently larger.

(b) Qualitative: There is a place on the Teacher Evaluation Sheet for comments. The following is a composite, largely in the students' own words, of both favorable and unfavorable comments in each class:

Unfavorable comments made by the discussion class.—"We should have been warned at registration that this was going to be a discussion class. This shows the drawbacks of modern education. Students don't know what they want, yet they run the class."

"We don't have enough background to know when we are right or wrong. Instead of giving us this background, the instructor just stands in front of the class, and leaves us up in the air, confused, in a fog, with nothing but a bunch of mixed-up facts and no concrete ideas. Too much time is taken in discussing procedure rather than subject matter; the discussions were irrelevant and uninteresting. We learned nothing from them. The whole thing is a farce, a waste of time and money, and a poor preparation for future psychology courses. We can get as much out of not coming to class. The instructor is failing to accept his responsibilities as a teacher."

"The text assignments are too long, the text is inappropriate for this kind of course, the type of test given is invalid for the discussion method, and some students who either don't attend class or don't participate still get the best grades, because no credit is given for class participation."

"This method won't work unless we are brought up in this system and are used to it, and unless everyone coöperates. It allows too much independent thinking."

"The instructor makes me feel like I'm being psychoanalyzed; he's probably writing a book, and we're his guinea pigs."

Two comments made during class discussions should also be added here:

"We work hard for our money to pay tuition here, and you're cheating us out of that money. The college pays you a salary to teach us, and you're failing to do what you're paid to do."

"You preach democracy to us, and then you don't practice it. You impose a method of teaching on us that we don't like—we came here to learn psychology, not democracy—and then when we vote that you should lecture to us, you refuse."

Unfavorable comments made by the lecture class.—"The instructor seldom smiles, and seems unhappy; he is unsocial, withdrawn, impersonal, difficult to know, so objective that one never gets to know his opinions, lacking in self-confidence, and unwilling to assume the leadership of the class, being too lenient in dealing with noisy students, too democratic. He doesn't exercise enough discipline, and is inclined to allow unnecessary discussion. He should challenge students when they say something, and should present his own views more; we don't know when we're accurate."

"Sometimes the lectures and other materials are not well organized, and do not clarify what is in the text or what is on the test."

(And, by a male:) "He shows sex discrimination."

Favorable comments made by the discussion class.—"The class was democratic, stimulating us to think on our own. This is a stimulating change from the class lectures I usually sleep through, and, in place of my usual apathy, makes me think. The students teach themselves."

"Coöperative student-teacher planning of the course is a good idea. It's only right that students have something to say about the way the class is run. We take up topics we choose ourselves, and I like the freedom from restraint."

Favorable comments made by the lecture class.—"The instructor is informative, presenting an abundance of material so that it can be readily absorbed. He covers the material well, explaining it clearly, without going in circles, and makes it interesting. He is well-versed in his subject, bringing in the latest developments. The examinations are fair."

"The frequent breaks make for a relaxed atmosphere; this was more democratic than most my classes."

"The course has helped me a great deal; I have developed a deeper insight into people as individuals through this course."

"The instructor is conscientious, going out of his way to help students. He is unusually sensitive to interpersonal relationships, and makes you feel that he is talking to you, that he is interested in you."

And, by a female: "Perhaps it's because he is young, but I feel very much at home around this instructor. I'm in favor of younger, more modern and informal instructors. . . ."

(c) Classroom behavior: In the discussion class various reactions to frustration were noted. Much aggressive behavior was expressed, usually verbally, occasionally in other ways. On one occasion, when the instructor asked the class what they would like to do with the time remaining in the class period, one student got to his feet, announced: "I suggest I leave and go study my physics," and marched out, slamming the door. An indication of the tone which sometimes prevailed, especially during the bimonthly gripe sessions, may be seen in comments offered after a particularly traumatic (from the instructor's view) session: "How can you just stand there and take it?" "Boy, I'd hate to be in your boots." At no time did the instructor notice aggression expressed by students towards students.

Other frustration reactions also occurred. One student declined to join a group. Others would not participate in their group's discussion, but held aloof, sometimes read.

Whenever the students were offered the opportunity to make suggestions for improving the conduct of the course, within the limitation that the instructor not lecture, the most frequent suggestion made was that the instructor lecture. Students who called attention to this fixated behavior termed it 'dependency'.

No corresponding behavior was noticed in the lecture class. When criticisms or suggestions were invited, minor suggestions—e.g., about taking a break during the two-hour period, or about the instructor's handwriting—were offered.

CONCLUSIONS

Achievement.—The method of teaching did not affect achievement on the type of examinations currently used in the course.

1) The results do not support the view, expressed by the students, that an instructor is failing in his duty to the students and in his responsibilities to the college if he merely leads discussions and does not give information. The results do not support the students'

view that, for effective learning to occur, they need explanation from the instructor. As Faw remarks: "There is certainly no reason to believe, judging from the test results, that the student-centered group was handicapped intellectually. The widespread observation on the part of the students that the student-centered group discussions resulted in less of an accumulation of facts and information was not borne by the facts. . . ." (2, p. 108).

2) In other words, the results do not support the view that learning is more effective when an instructor speaks, outlines, and illustrates with movies and demonstrations, than when he contributes only an opportunity for the students to discuss the material—or not even this: During the third two-week period, the discussion bore no relation to the text or test, but the discussion class did as well as the lecture class. Apparently the instructor's explanations and demonstrations were a difference which made no difference.

3) The students' claim that they lack the background necessary for a creative discussion is not supported. As Faw states: "Actually the discussion brought out a wealth of facts and information which were authentic and verified. In many instances the instructor felt that the presentation of facts by students was as comprehensive and was much more interesting than if he had made the presentation" (3, p. 109).

Some degree of validation for this view may be seen in the fact that, when the discussion material was presented to the lecture class, it was favorably received as meaty and relevant. This favorable reception was reflected both in the class response at the time and on the Teacher Evaluation Sheets.

4) On the other hand, the results do not support the view, held by the writer, that giving the students the opportunity to think and to express their ideas will make for better learning than will forcing the students into a more passive, receptive rôle, in which they memorize facts. On their own discussion material, the discussion class was not consistently better than the lecture class. Our learning principles say that recitation is more effective than memorization, but the principle seems not applicable to this classroom situation.

5) One would expect that throwing the students on their own might stimulate some and frustrate others, and that this would cause a greater spread of achievement scores among the students in the discussion class. This did not occur.

ATTITUDE

The method of teaching did affect attitudes towards the course.

1) There is a difference between one's achievement and one's feeling of achievement. Faw also found this "discrepancy between what students feel about their accumulation of knowledge and actual test results" (2, p. 109). The discussion class felt that they had learned little; the lecture class felt that they had learned a great deal. As measured by examinations, both learned an equal amount.

2) Students' feeling of satisfaction, of liking or disliking a course, is a function of methodology or technique of teaching, as well as of the instructor's personality and the course content. This difference makes a difference.

3) One would expect that throwing the students on their own might stimulate some and frustrate others, and that this would cause a greater spread of attitude scores among the students in the discussion class. This occurred.

4) Even apart from the difference in the degree to which each class liked the course, there was an important difference in the type of reaction of the two classes. In the lecture class, the students took an evaluative attitude towards the person of the instructor, commenting both favorably and unfavorably upon him. In the discussion class, the students seemed simply to accept the instructor, and responded largely to the method of teaching. In other words, in the discussion class rejection of the teaching technique was not accompanied by rejection of the instructor as a person, and enthusiasm for the teaching technique was not accompanied by any transference of this positive feeling to the person of the instructor.

DEMOCRATIC AND AUTOCRATIC LEADERSHIP IN THE CLASSROOM

1) While the psychologist may define democracy and autocracy in terms of whether the decision-making function lies with the group or remains with the leader, the students are likely to define these concepts by other criteria:

(a) Are we pleased? The group-discussion class was perceived by the students as an attempt on the part of the instructor to force something upon them. The lecture class was perceived by the students as being both factual and entertaining, as serving their needs in such matters as breaks during the lecture period and generosity with grades, and as meeting their expectations.

(b) What kind of attitude does the instructor have? The students perceive as democratic any instructor who allows them a certain amount of freedom in such matters as reading newspapers or whispering while in class, who is courteous, relaxed, humorous, who states things in a conditional, non-authoritarian manner. "He has coffee with students, and learns their names." "He's willing to take time to chat with them after class." "His decisions are based upon what he believes is best for them, rather than upon what is most convenient for himself."

This was the perception of the students in the lecture class. In the discussion class, the instructor never used the term 'democratic'. Nevertheless, from the first week the students spontaneously referred to the class as democratic. Having so defined the class, they proceeded to:

(a) Project this definition onto the instructor, i.e., they insisted that he had defined the class as democratic, and that he had been preaching democracy to them;

(b) Chide him, whenever he refused to yield to any student demand, with failing to live up to his preachings.

2. There is no evidence here that students prefer a group-discussion and group-decision method, or that they like integration and resent domination. Even in the vital matter of grades, the class who decided for themselves the proportion of A's and B's were less satisfied than was the class in which the instructor made the decision.

3) Defining type of leadership in terms of decision-making, we find no evidence here that autocracy produces more aggression than does democracy.

4) Finally, there is no evidence here that a leader best implements the type of instruction in which he believes (4, p. 425). This instructor, while enjoying both methods of teaching, strongly prefers the group-discussion method of leadership, whether in the classroom, in industry, in hospital and prison administration, or in any other situation. He finds the client-centered philosophy, the non-directive technique, congenial. Nevertheless, he apparently does a better job as a directive leader.

This experiment was carried out during the Winter Quarter. It was repeated, with essentially similar results, during the Spring Quarter. During the Summer Session following these two quarters, the writer mailed to all four groups of students an objective exami-

nation, an essay examination, and a course evaluation scale. The objective examination was also given to the Summer Session students. The proportion of returns from each of the four post-final sections was approximately one-third. A comparison of Winter, Spring, and Summer students failed to show any evidence of post-final forgetting. Equality of achievement between lecture and discussion sections remained. However, when the course was over and grades were in, the two types of class, discussion and lecture, showed an equally favorable attitude towards the course.

SUMMARY

Keeping the emotional climate similar in the two classes, an instructor taught one elementary psychology class by group-discussion; the other by lecture. The group-decisions were used in order to determine the conduct of the course in such matters as examinations and grades, and also to cover the course content. Following the lead of Lippitt and White, the instructor imposed the same decisions on the control class, and covered the same material by means of lectures and demonstrations.

Achievement on course content was not significantly different. The morale of the lecture class, however, was significantly higher than that of the 'democratic' class.

REFERENCES

- 1) H. H. Anderson. "Domination and socially integrative behavior." In *Child Behavior and Development*, (Ed., R. G. Barker, et al.). New York: McGraw-Hill, 1943, 459-483.
- 2) V. E. Faw, "A psychotherapeutic method of teaching psychology." *Amer. Psychol.*, 1949, 4, 104-109.
- 3) R. Lippitt and R. K. White. "The 'social climate' of children's groups." In *Child Behavior and Development* (Ed., R. G. Barker, et al.). New York: McGraw-Hill, 1943, 485-508.
- 4) C. R. Rogers. *Client-centered Therapy*. Boston: Houghton Mifflin, 1951.

STUDENT ATTITUDE AND ACHIEVEMENT IN A COURSE IN INTRODUCTORY STATISTICS

A. W. BENDIG and J. B. HUGHES III¹

University of Pittsburgh

In faculty conversations concerning the teaching of psychology, the viewpoint is often expressed that instructors of courses in introductory statistics are faced with the special problem of the emotional attitude of the student toward the course. A large number of students seem to fear that they will be unable to learn the content matter of the course, either because of an inadequate background in mathematics or because of an emotional hangover acquired from previous somewhat traumatic exposures to high school and college mathematics courses. If this problem really exists, it has failed to be recognized in the elaborate recommendations about undergraduate training in statistics that have been sponsored and reported by the American Statistical Association (3, 12, 13, 14). Attention is paid to course content and structuring, and to the required educational background of both instructors and students. However, little attention is directed toward the effect of the mental attitude of the student in the biological and social sciences, who is often required to take such a course as part of his undergraduate training (1), upon his academic achievement in statistics. The results of Margaret and Sherriffs (9), indicating that the attitude of students toward psychology is materially related to achievement in the introductory psychology course, suggest that a similar relationship may be present in the introductory course in statistics.

The problem seems to be threefold: (a) can such an attitude be reliably measured, (b) if so, is this attitude related to student achievement in the course, and (c) what are the significant concomitants of such an attitude. The purpose of this paper is to report the results of a preliminary investigation of these questions.

THE ATTITUDE SCALE

In the Fall, 1950-51, semester one class period of the senior author's course in introductory psychological statistics was devoted

¹ The authors wish to express their appreciation to Dr. Henry Zagorski and Mr. H. J. Maginley for their coöperation in obtaining a part of the data for this study.

to a discussion of student attitude toward the course. The instructor presented to the students his observations and hypotheses concerning student attitude and course achievement and encouraged verbalizations by the students of their fears about what would be required of them in the course. The students were asked to write several short paragraphs on their attitude and anonymously turn them in to the instructor at the next class period. Parenthetically, it might be noted that the reaction of the students was both enthusiastic and helpful. Following the collection of the paragraphs, fifty negatively oriented statements were derived from the paragraphs covering all aspects of the attitudes reported by the students. For example, three of the statements were:

"I have less math ability than most of my friends."

"I'll probably never use what I learn in this course after I graduate."

"My grade in this course will probably be closer to a D than to B."

These fifty statements were mimeographed as a test booklet with instructions requesting the students to rate each of the statements on a five-point scale as to how well the statement reflected their feelings about the course. This test was administered to all students in introductory psychological statistics classes at the beginning of the following semester: Spring 1950-51. In addition, the letter grade received by each student was obtained at the end of that semester. A total of fifty students took both the attitude test at the beginning of the semester and received a completed letter grade at the end of the course.

The first step in the analysis of the data was to dichotomize each item. The fifty items on the original form of the test were analyzed to find the number of students selecting each of the five rating choices on each item. Each item was then dichotomized by combining adjacent response categories to achieve, as closely as was possible, a cutting point that would divide the respondents on that item into two equally sized groups. On most of the items approximately half of the subjects chose the first, most favorable choice with the remaining subjects spread over the other four, less favorable ratings. The combined favorable ratings for each item were given a weight of one, the combined unfavorable categories a weight of zero. A scoring key for the combined favorable choices on the fifty items was made and each of the answer sheets scored, each

subject receiving as a raw score the number of keyed favorable responses that he made. These raw scores ranged from four to forty-eight out of a possible fifty. Each item was then correlated with the total score on the test to get estimates of item reliability. These item reliabilities ranged from 0.07 to 0.87.

An estimate of the validity of each item as a predictor of achievement was obtained by correlating the item with students' letter grades in the statistics course. The five letter grades were dichotomized by combining the A and B students into one group, and the C, D, and F students into a second group. The letter grade criterion used here was found to have a reliability of 0.73 in a previous investigation (1). Phi correlation coefficients were computed between the dichotomized letter grade criterion and each item. These correlations ranged from -0.17 to 0.44 . The two correlation coefficients for each item (reliability and validity) were each weighted by the standard deviation of the item, graphically plotted, and the thirty items having the highest reliability and validity selected for the final form of the test, using the selection technique outlined by Gulliksen (7, pp. 382-385). A revised test key containing these thirty items was used to rescore the answer sheets of the original group of fifty students (Sample A) and the results are reported in Table 1. The reliability of the new key, using Kuder-Richardson Formula 20 (7, p. 224), was found to be 0.96. Raw scores from the new key were correlated (by the usual product-moment method) with quantified letter grades (A being assigned a numerical value of 4, B a value of 3, etc.) and a correlation of 0.24 was found. Such a magnitude of correlation indicates that approximately five per cent of student achievement in Sample A could be accounted for by variation in student attitude as measured by our test.

Cureton (2) has amusingly pointed out the fallacy in validating

TABLE 1.—RESULTS OF TWO SAMPLES OF STUDENTS ON THE STATISTICS COURSE ATTITUDE SCALE

	Original Sample	Validation Sample
Number of Students	50	71
Mean	17.48	19.51
Standard Deviation	9.54	6.73
Reliability (KR-20)	0.96	0.90
Validity	0.24	0.21

a new test upon the group from which the test is derived. To check the reliability and validity of the new thirty-item attitude scale, this test was given to all students in introductory psychological statistics classes at the beginning of the Fall, 1951-52, semester. In addition, each student was given the Kuder Preference Record (8) and a biographical information blank. The data from these last two instruments will be reported in a later section of this paper. A total of seventy-one students completed all three measuring instruments and also received a complete achievement grade for the course. The results on the attitude scale for this group (Sample B) can also be found in Table 1. The reliability of the test (again using Kuder-Richardson Formula 20) was 0.90 and the product-moment correlation between raw scores on the test and quantified letter grades was 0.21. The differences between the pairs of reliability and validity coefficients were tested for statistical significance by the *r*-to-*z* technique outlined by Edwards (4, pp. 131-132). The drop in the reliability of the scale in Sample B (0.06) was significant at the .05 level of confidence ($t = 2.49$), while the decrease in the validity coefficient (0.03) was not significant ($t = 0.17$). While scale reliability was significantly lower in Sample B, its magnitude (0.90) argues that the thirty-item attitude test remains a reliable instrument in the new sample. The similarity in validity coefficients in the two samples leads to the conclusion that there is a consistent relation between student attitude and achievement in introductory statistics, with individual student differences in attitude accounting for four to five per cent of their variability in achievement.

CONCOMITANTS OF ATTITUDE

As mentioned above, the seventy-one students in Sample B had been given, in addition to the attitude test, the Kuder Preference Record and a biographical data sheet. The raw scores from the nine interest scales on the Kuder were converted to stanine scores, based upon undergraduate norms of University of Pittsburgh students as prepared by the University Testing Service. From the personal data supplied by the students on the biographical information sheet six variables that were hypothesized as possible concomitants of attitude were derived and subsequently dichotomized for convenience in handling. These variables were:

Major subject.—Students majoring in psychology were given a numerical weight of one; non-majors were weighted zero.

Sex.—Male students were weighted one: female students zero.

Amount of high school mathematics.—Students taking 2.5 or more years of mathematics in high school were weighted one: those with two years or less were weighted zero.

Amount of college mathematics.—Those having taken one or more semesters of college mathematics courses weighted one: students not having had any college mathematics training were weighted zero.

Number of previous psychology courses.—Three or more semesters of psychology was weighted one: one or two semesters weighted zero.

College class.—Senior, graduate, and non-degree students were grouped and weighted one: sophomore and junior level students were given a weight of zero.

The Kuder stanine scores were correlated with raw scores on the attitude scale by the conventional product-moment method. Point-biserial correlations (6, p. 329) were computed between the continuous scores on the attitude scale and the dichotomous Major Subject and Sex variables. The correlations between attitude scores and the remaining four variables were computed using the biserial correlation formula (6, p. 324), since a dichotomy had been arbitrarily imposed upon these continuous variables. These correlations are reported in Table 2. Four of the fifteen variables demonstrated relationships with attitude that were significant at the .01 level of confidence. The Kuder Computational scale, amount of high school mathematics, and amount of college mathematics were all positively related to a favorable attitude toward the course. Number of psychology courses was negatively related to a favorable attitude.

The final question asked of the data was what was the minimum number of variables that would best predict student attitude and how much did each of these variables contribute to total attitude variation. The simplified technique reported by Gengerelli (5) was used to select the five most important predictors of attitude out of the total group of fifteen variables. The five selected included the four variables reported above as showing significant covariation with attitude, plus the Kuder Persuasive scale whose correlation with the attitude scale (.22) fell just short of statistical significance (.23 is significant at the .05 level with seventy degrees of freedom). From the correlations of these five variables with the attitude criterion and the intercorrelations of the variables, multiple regres-

TABLE 2.—CORRELATIONS OF FIFTEEN VARIABLES WITH SCORES ON THE STATISTICS COURSE ATTITUDE SCALE (N = 71)

Variable	Correlation	Beta Weight	Per cent of Criterion
Kuder Preference Record			
Mechanical	0.16		
Computational	0.31*	0.32	10
Scientific	0.20		
Persuasive	0.22	0.61	13
Artistic	0.02		
Literary	0.03		
Musical	0.08		
Social Service	0.07		
Clerical	-0.04		
Major Subject	-0.21		
Sex	0.13		
Amount of High School Math	0.39*	0.30	12
Amount of College Math	0.40*	0.28	11
Number of Psychology Courses	-0.42*	-0.54	23
College Class	0.15		
Total.....			69

* Significant at the 0.01 level.

sion coefficients were computed using the method detailed by Peters and Van Voorhis (11, pp. 226-229). These coefficients are reported as Beta weights in Table 2. The multiple correlation (corrected for shrinkage) between these five predictors and the attitude criterion was .82. The final column in Table 2 gives the multiplicative product of each variable's correlation with attitude and its Beta weight. Guilford (6, p. 432) suggests that variable weights, derived in this fashion, indicate the percentage of total criterion variation that is accounted for by each predictor variable and that the sum of these weights (actually the squared multiple correlation coefficient) gives the percentage of criterion variability predicted by the entire battery of predictors. The weights found in Table 2 show that Kuder Computational and Persuasive scales, amount of high school mathematics, and amount of college mathematics, each accounts for approximately twelve per cent of attitude variability, while number of psychology courses predicts almost twice the variation (twenty-three per cent) represented by any other single pre-

dicator. The battery of five variables shows covariation with sixty-nine per cent of the criterion variability.

DISCUSSION

In terms of the original three questions proposed at the beginning of this paper we now have some tentative answers. A thirty-item test was constructed to measure students' attitudes toward an introductory statistics course designed for students majoring in psychology and related biological and social sciences. This test showed a satisfactory magnitude of reliability (.90) when given to a new sample of seventy-one students. The empirical validity of the test in actually measuring 'true' student attitude is as yet undetermined, but the rational or face validity inherent in the technique by which it was constructed suggests that such an attitude dimension is actually being tapped by the test.

Student attitude measured by the test has been shown in two samples to account for four to five per cent of student variability in course achievement. Theoretically, such a finding corroborates the hypothesis voiced by instructors of introductory statistics courses that student attitude is a partial determinant of student achievement. Practically, this finding suggests two recommendations: (a) instructors of such courses should attempt to modify and allay the fears of their student early in the course if maximum student achievement is the academic goal, and (b) any battery of measuring instruments devised to predict achievement in similar courses should include some personality or temperament measuring device that can be logically expected to tap this attitude dimension. Such an attitude scale can be expected to predict a portion of the criterion variance not measured by the aptitude tests most commonly used in such prediction.

It seems quite logical that increased exposure to mathematics courses in high school and college, plus an interest in computational pursuits, should dispose a student toward a statistics course. Such a background gives the student a familiarity with, and a liking for intellectually handling, the quantitative concepts so important a part of statistics. In passing, it may be noted that Kuder (8, p. 5) lists the occupations of Statistician and Statistical Clerk as two showing major interests on the Computational scale. The relation of persuasive interests to attitude is less clear, but it can be pointed out that a course in statistics is only partially quantitative in con-

test. In addition, Kuder (8, p. 7) lists both Advertising Statistician and Business Statistician occupational groups as showing high interest on both the Computational and Persuasive scales. The somewhat surprising result reported in Table 2 is the strong negative relation between number of psychology courses and favorableness of attitude. Such a finding may be explained by the type of student enrolling for the course. Most of the students in this course who have taken more than two semesters of psychology are undergraduate psychology majors (the correlation between Major Subject and Number of Psychology Courses was .56, the highest intercorrelation found between any of the fifteen variables) who are required to take this course as one of the departmental prerequisites for graduation. The students with few previous psychology courses are typically majors in some other field who take the course as an elective subject. Since the graduation and possible later admission to graduate work of the student who is majoring in psychology is partially, at least, dependent upon his achievement in this course, we might logically expect him to enter the course with more fear and trembling than the non-major elective student. Another possible explanation (one which the authors indignantly reject) is that our psychology majors are refugees from quantitative thinking and demonstrate their antipathy toward a forced exposure to such training by a negative attitude toward the course. The first explanation offered seems much more plausible (and ego-placating) to us than does the second.

REFERENCES

- 1) A. W. Bendig. *The Reliability of Letter Grades in a Course in Introductory Psychological Statistics*. (Mimeographed report) Department of Psychology, University of Pittsburgh, Pittsburgh, Pa., 1952.
- 2) E. E. Cureton. "Validity, reliability, and baloney." *Educational and Psychological Measurement*, x (1950) 94-96.
- 3) S. Dutka and F. Kafka. "Statistical training below the college level." *American Statistician*, iv (1950) 6-7.
- 4) A. L. Edwards. *Experimental Design in Psychological Research*. New York: Rinehart, 1950.
- 5) J. A. Gengerelli. "A Simplified method for approximating regression coefficients." *Psychometrika*, xiii (1948) 135-146.
- 6) J. P. Guilford. *Fundamental Statistics in Psychology and Education* (2nd ed.). New York: McGraw-Hill, 1950.
- 7) H. Gulliksen. *Theory of Mental Tests*. New York: Wiley, 1951.
- 8) F. G. Kuder. *Revised Manual for the Kuder Preference Record*. Chicago: Science Research Associates, 1946.

9) A. Margaret and A. C. Sherriffs. "Personal factors influencing the learning of the first course in psychology." *Journal of General Psychology*, XXXVII (1947) 67-77.

10) M. D. Munn and G. H. Schlanch. "A study of the content of a course in elementary statistics." *American Sociological Review*, XIV (1949) 679-681.

11) C. C. Peters and W. R. Van Voorhis. *Statistical Procedures and Their Mathematical Bases*. New York: McGraw-Hill, 1940.

12) G. W. Snedecor. "A proposed basic course in statistics." *Journal of the American Statistical Association*, XLIII (1948) 53-60.

13) H. M. Walker. "Statistical literacy in the social sciences." *The American Statistician*, v (1951) 6-12.

14) S. S. Wilks. "Undergraduate statistical education." *Journal of the American Statistical Association*, XLVI (1951) 1-18.

DETERMINING STUDENTS' CONCEPTS OF EFFECTIVE TEACHING FROM THEIR RATINGS OF INSTRUCTORS

WILLIAM E. COFFMAN¹

Educational Testing Service

During the twenty-seven years since the use of student ratings of college instructors first appeared in the literature, discussion has centered around two viewpoints. Critics of the use of student ratings have pointed to the inadequacy of the student as a judge of teaching ability, emphasizing his lack of experience and the difficulty of reporting judgments which are free from subjective bias. Proponents of student ratings, on the other hand, have tended to emphasize that effective learning results from the interaction of student and teacher and that however biased ratings may be, they are valuable as a source of information concerning student reactions to the behavior of teachers.

Meanwhile evidence has accumulated concerning students' ability to rate instructors, and more and more institutions have turned to student ratings for one purpose or another. (5) Well designed studies have demonstrated that student ratings are reliable. (6) Others have provided evidence that ratings on multiple attributes need not be influenced by halo. (3) At least one study has shown that student ratings are similar to ratings of alumni after ten years of seasoning. (2)

But neither party to the controversy is content with demonstrations of reliability, absence of halo, and conformity to judgments of mature individuals who are not teachers. Those who question students' ability to rate their instructors require proof that student values are sound before accepting the validity of ratings; those who focus on understanding students need the same information as a basis for interpreting ratings.

¹ This report covers one aspect of a general investigation of Student Rating of Instruction conducted under a grant of the Research Foundation of Oklahoma A. and M. College. The writer wishes to acknowledge his debt to the instructors and students who wholeheartedly cooperated in the investigation and to Educational Testing Service for help in preparing the final report.

A recent study conducted at Brooklyn College (7) approached the problem by asking students directly to select from a list of ten the three items which they thought most important to good teaching. The method raises certain questions. Did these ten items cover all of the aspects of good teaching which the students might value? Were students influenced in their responses by cultural stereotypes concerning the nature of college teaching?

An opportunity to approach the problem of student values in a different way presented itself when approximately two thousand students at Oklahoma A. and M. College rated fifty-five teachers as part of a comprehensive investigation of the *Oklahoma A. and M. Rating Scale for Instructors*. The Rating Scale consists of eighteen specific traits and a general estimate of the instructor's effectiveness. For each rating the student is asked to check a graphic scale upon which are placed five descriptive phrases. The traits and descriptive phrases are:

1) Preparation for Class Meetings: Class meetings very carefully planned. Usually well prepared. Preparation often inadequate. Little preparation. No preparation.

2) Teacher's Interest and Enthusiasm in Subject: Very enthusiastic and interested. Frequently shows enthusiasm. Only mildly interested. Very seldom shows enthusiasm. Subject irksome to him.

3) Ability to Arouse Interest in Students: Interest usually runs high. Students frequently show interest. Students occasionally show interest. Students seldom interested. Majority inattentive most of period.

4) Organization of Course: Every lesson well organized. Most lessons well organized. Some organization but not always clear. Very little organization. No organization.

5) Thinking Demanded of Students: Thinking always necessary. Work demands much thinking. Some thinking required. Very little thinking required. No thinking required.

6) Assignments: Clear and definite. Carefully given but indefinite. Definite; often hurriedly given. Rather indefinite and often hurriedly given. Very indefinite; usually hurriedly given.

7) Sense of Proportion: Stresses fundamental topics, disregards trivial details. Spends most of time on important topics, stresses few details. Stresses important topics and details equally. Spends more time on details than on important topics. Often neglects subject for other irrelevant topics.

8) **Enunciation:** Speaks very clearly and distinctly. Distinct but not loud enough at times. Loud enough but not distinct. Words sometimes indistinct and not easy to hear. Words very indistinct often impossible to hear.

9) **Sense of Humor:** Has keen sense of humor. Frequently shows real humor. Humor occasionally, but not often exhibited. Manifests little or no humor. Humor obviously not spontaneous.

10) **Scholarship:** Knowledge of subject broad and accurate. Knows appreciably more than is in text. Knowledge limited to text material. Knowledge apparently deficient at times. Knowledge very plainly deficient.

11) **Ability to Express Thought:** Words come easily; meaning always clear. Some hesitation; meaning always clear. Some hesitation for words; meaning at times not clear. Much hesitation for words; meaning often not clear. Meaning almost never clear.

12) **Feeling between Instructor and Students:** Complete harmony. Feeling of good-will prevails. Neither good-will nor antagonism seems to prevail. Students frequently antagonized. Instructor tends to antagonize class.

13) **Prevalence of Cheating on Examinations:** By no one. Only infrequently. Occasionally noted. By many students. By most students.

14) **Self-confidence:** Sure of himself; meets difficulties with poise. Seldom if ever disconcerted. Fairly self-confident; occasionally disconcerted. Often confused by students. Hesitant, timid, uncertain.

15) **Tolerance and Liberality:** Invites differences of opinion. Welcomes differences of opinion. Shows no personal response to opposition. Sometimes impatient when students oppose his views. Easily aroused to temper by opposition.

16) **Punctuality in Meeting and Dismissing Classes:** Prompt. Rarely late. Occasionally late. Often late. Very often late.

17) **Personal Appearance:** Well-groomed; clothes neat, clean, in good taste. Usually tidy and careful as to appearance. Sometimes careless as to appearance. Usually rather untidy and careless as to appearance. Slovenly; clothes and person untidy.

18) **Personal Peculiarities:** Free from mannerisms. Free from annoying mannerisms. Annoying mannerisms not serious. Annoying mannerisms often detract. Constantly exhibits annoying mannerisms.

19) General Estimate of Instructor as a Teacher: Excellent. Good. Fair. Poor. Very poor.

With ratings available for eighteen specific traits as well as for a general estimate of the instructor's effectiveness, it was possible to calculate a matrix of intercorrelations and to study through factor analysis the question: What factors are reflected in the over-all ratings of these students? In other words, this study reports an attempt to determine students' concepts of effective teaching from their ratings of instructors.

PROCEDURE

The raw data for the study were the average ratings of each of the fifty-five instructors on each of the nineteen scales. Coefficients of internal consistency had been computed using the method suggested by Horst. (4) The intercorrelations of the nineteen variables were calculated by the product-moment method. Four reference factors were extracted by the complete centroid method, (8) and then rotated to simple structure using the single-plane method of rotation. (8) The reference factor matrix, the communalities, and the coefficients of internal consistency are given in Table I. The final rotated matrix of factor loadings is given in Table II. Table III contains the direction cosines of the final reference vectors, and Table IV, their intercorrelations. It can be seen that the final reference vectors are almost orthogonal; therefore, it seemed unnecessary to compute the direction cosines of the primary reference vectors, their intercorrelations, and the projections of the variables on the primaries. Interpretations are made directly from Table II. The correlations which seem especially significant for naming the factors are starred.

RESULTS

Factor A appears to be a factor of empathy. There are high correlations with (1) ability to arouse interest in students, (2) sense of humor, (3) feeling between instructor and student, (4) tolerance and liberality, and (5) general estimate of instructor as a teacher.

The factor seems to include more than a tribute to simple friendliness or to the instructor's tendency to be easy. Here is student recognition that good teaching often requires an awareness of problems learners encounter, a willingness to be patient with student

TABLE I.—THE CENTROID FACTOR LOADINGS

Scale No.	I	II	III	IV	h ²	r _{ii} [*]
1	.60	-.52	-.43	.22	.86	.90
2	.49	.17	-.37	.18	.44	.70
3	.78	.37	.08	.13	.77	.90
4	.72	-.42	-.37	.25	.89	.84
5	.09	.32	-.31	.07	.21	.81
6	.41	-.31	.23	.12	.33	.86
7	.62	.25	-.04	.12	.46	.79
8	.54	.06	.09	-.31	.40	.86
9	.58	.32	.31	.11	.55	.91
10	.31	.10	-.49	.11	.36	.69
11	.63	-.16	-.09	-.43	.62	.89
12	.71	.26	.43	.14	.78	.92
13	.38	.14	-.13	-.11	.19	.66
14	.67	.10	-.32	-.25	.62	.76
15	.58	.16	.46	.10	.58	.86
16	.24	-.68	.20	.10	.57	.96
17	.27	-.46	.48	-.16	.54	.92
18	.53	.13	.29	-.30	.47	.86
19	.91	.16	.03	.18	.89	.84

* Calculated by Horst Formula (*Psychometrika* March 1949).

limitations, an appreciation of the world of the student as well as of the world of the scholar.

The relatively high value placed on empathy by these students is not necessarily representative of the situation generally. In the Brooklyn study (7) 'sympathetic attitude' was ranked relatively low. The difference noted here may result from the difference in approach; students asked directly may be hesitant to admit that they feel attitude is important. However, the difference can also be explained in terms of differences in academic aptitude in the two groups. The more difficulty students have with class work, the more they are likely to value teachers who are helpful and understanding.

Factor B is identified as a factor of organization. There are high correlations with: (1) preparation for class meetings and (2) organ-

TABLE II.—THE OBLIQUE FACTOR MATRIX AFTER ROTATION

Scale No.	Scale Name	A	B	C	D
1	Preparation for Class Meetings	-.02	.92*	.02	-.05
2	Teacher's Interest and Enthusiasm in Subject	.22	.40	-.34*	.27
3	Ability to Arouse Interest in Students	.65*	.17	-.06	.53*
4	Organization of Course	.13	.89*	.01	.04
5	Thinking Demanded of Student	.05	.04	-.42*	.19
6	Assignments (definiteness of)	.26	.31	.38	-.03
7	Sense of Proportion	.46	.23	-.10	.40
8	Enunciation	.17	.15	.23	.52*
9	Sense of Humor	.64*	-.03	.11	.39
10	Scholarship	.00	.40	-.40*	.19
11	Ability to Express Thought	-.03	.41	.29	.55*
12	Feeling Between Instructor and Student	.77*	.01	.25	.40
13	Prevalence of Cheating on Examinations	.12	.17	-.07	.36
14	Self-Confidence	.07	.43	-.08	.61*
15	Tolerance and Liberality	.65*	-.02	.32	.30
16	Punctuality in Meeting and Dismissing Classes	.00	.45	.54*	-.30
17	Personal Appearance	.11	.13	.72*	-.02
18	(Lack of) Personal Peculiarities	.31	-.01	.34	.53*
19	General Estimate of Instructor as a Teacher	.63*	.41*	.03	.46*

ization of the course, and a relatively high correlation with the overall rating. The positive correlations with teacher's interest and enthusiasm in subject, scholarship, ability to express thought, self-confidence, and punctuality can be accepted as reflections of the

TABLE III.—THE DIRECTION COSINES OF THE FINAL REFERENCE VECTORS

	A	B	C	D
I	.50	.54	.18	.55
II	.41	-.61	-.57	.52
III	.53	-.55	.73	-.07
IV	.54	.18	-.34	-.65

TABLE IV.—THE INTERCORRELATIONS OF THE FINAL REFERENCE VECTORS

	A ^a	B	C	D
A	.99			
B	-.17	1.00		
C	.06	-.02	1.01	
D	.10	-.10	-.03	1.00

fact that people who have well organized courses and who are prepared for each class also tend to be enthusiastic about their subject, to express thoughts more clearly, to be more punctual, and to appear more scholarly and self-confident. The absence of correlation with 'thinking demanded of student' appears significant also. The students seem to value a particular type of organization, one where the instructor clearly indicates what is expected of the student and so structures the situation that they can 'dig in' without experiencing the 'agony of suspended judgment' or the chaos of planning an independent project. These students, then, value instruction which is characterized by carefully organized sequences and well prepared class sessions.

'Systematic organization of subject matter' was one of the qualities valued by the Brooklyn College students also. (7) The students placed it second for science and social science teachers, and fifth for humanities teachers.

Factor C seems to represent the punctual, neat, and 'normal' individual who is the opposite of the 'absent-minded professor'. There are high positive correlations with punctuality and personal appearance. There are negative correlations with teachers interest and enthusiasm in subject, thinking demanded of students, and scholarship. The person with a low score on this factor is a mild prototype of the absent-minded professor—too busy to get a needed haircut, arriving late to class occasionally because he has been pursuing some academic interest, engrossed in his subject, a scholar and a Socrates in the classroom.

This factor is not related to the over-all rating; students apparently describe instructors through their ratings and the factor appears in the analysis, but it is not reflected in their judgment of teaching competence. This suggests that students are not considering surface manifestations when they make their judgments of teaching effectiveness.

Factor D appears to be a factor of verbal fluency. There are moderately high correlations with: (1) ability to arouse interest in students, (2) enunciation, (3) ability to express thought, (4) self-confidence, and (5) lack of personal peculiarities, and with the overall rating. The correlations with enunciation and with ability to express thought are somewhat lower than one would expect if the factor were general speaking ability, and the correlations with the other scales seem about right if the factor is verbal fluency. Even the negative correlation with 'punctuality in meeting and dismissing classes' could be more than an unexplained irregularity in the data, since there is a tendency for the verbally fluent to keep talking after the bell marking the end of a period has rung.

The Brooklyn College students had no opportunity to judge verbal fluency. They rated 'good speaking ability' low (sixth for humanities teachers and seventh for social science and science teachers), but they placed higher value on 'ability to explain clearly' (first for science teachers and fourth for humanities and social science teachers).

DISCUSSION

The picture one gets of students from this analysis is not one of incompetent judges assessing teaching on the basis of superficial evidence. Who will argue that understanding of the learner, organization of the course, and verbal fluency are insignificant aspects of good teaching? And if the picture of the ideal teacher seems unduly restricted, shall we attribute it to the incompetence of students as judges? Certainly not until we have assessed the limitations of the rating scale and examined more fully the values of faculty members themselves.

Here is an example in point. These students appear at first glance to place little value on stimulation to thinking as an aspect of good teaching. At least the scale 'thinking demanded of students' is but slightly related to the general estimate of effectiveness. But the scale is somewhat ambiguous. Does it refer to the thinking which follows stimulating contact with the instructor—or does it mean the thinking required because the instructor left the student to shift for himself? Perhaps a more clearly described scale would bring different results.

On the other hand, the data from the Brooklyn study suggest that for certain courses students do not value stimulation to thought.

Only seventeen per cent of the students listed it as one of the three most desirable qualities for science teachers. Is it possible that teachers themselves have emphasized other aspects of the teaching process so much that students have failed to appreciate the values in stimulation to thought?

By revising the scale and repeating the analysis one might clarify the interpretation. Three possible types of revision seem indicated. (1) Some of the categories in the present scale have low reliability and may be discarded without significant loss of information—(teacher's interest and enthusiasm in subject, scholarship, prevalence of cheating on examinations). If evidence relative to these is desired, other means of collecting information should prove more reliable. (2) Factors A and D are definitive enough to be included as separate scales in the rating sheet as a check on the accuracy of the interpretation given here. (Factor B appears to be clearly described in Scale 1 and Scale 4.) (3) One or more scales related to an hypothecated stimulation to thinking factor might be added in place of the ambiguous Scale 5 as a means of further clarifying the position of students relative to the importance of this aspect of teaching.

The factor analysis approach might also be used to test for changes in student values as a result of faculty effort. One could analyze ratings previous to such effort and then repeat the analysis following a period of special emphasis on helping students understand the goals of instruction. If, for example, a preliminary analysis revealed that students placed little value on stimulation to thinking as an aspect of good teaching, and if a group of instructors directed their efforts toward helping students understand and accept stimulation to thinking as a legitimate goal of instruction, a subsequent analysis should result in the appearance of a factor of stimulation to thinking with significant loadings for the over-all rating.

REFERENCES

- 1) P. C. Baker and H. H. Remmers. "The measurement of teacher characteristics and the prediction of teaching efficiency in college," *Review of Educational Research*, 1952, 22, 224-227.
- 2) A. J. Drucker and H. H. Remmers. "Do alumni and students differ in their attitudes toward instructors?" *Journal of Educational Psychology* 1951, 42, 129-143.
- 3) R. L. Hobson. "Some psychological dimensions of academic adminis-

trators" *Studies in Higher Education* LXXIII, Lafayette, Indiana: The Division of Educational Reference, Purdue University, 1950.

4) P. Horst. "A generalized expression for the reliability of measures," *Psychometrika*, 1949, 14, 21-32.

5) F. J. Mueller. "Trends in student ratings of faculty," *Bulletin of the American Association of University Professors*, 1951, 37, 319-324.

6) H. H. Remmers. "Appraisal of college teaching through ratings and student opinion," in *The Study of College Instruction* (Fowler D. Brooks, Editor) Chicago: University of Chicago Press, 1939, 227-240. Twenty-seventh Yearbook, National Society of College Teachers of Education.

7) J. W. Riley, Jr., B. F. Ryan, and M. Lifshitz. *The Student Looks at His teacher*, New Brunswick, N. J.: Rutgers University Press, 1950.

8) L. L. Thurstone. *Multiple Factor Analysis*, Chicago: University of Chicago Press, 1947.

EFFECT OF SLANTED TEXT UPON THE READABILITY OF PRINT¹

MILES A. TINKER

University of Minnesota

When printed copy is lying flat, the accommodation changes required for clear seeing as the eyes shift from fixation to fixation along a line of print are relatively slight and do not appear to interfere with the mechanics of reading. Furthermore, on flat copy the word forms are clearly visible. In some books the printed material does not lie flat, i.e., as in many thick texts, dictionaries, and bound journals. In these books there tends to be considerable curvature of the page near the inner or gutter margin where the pages are bound. To what degree does this condition interfere with easy and rapid perception in reading? Apparently there are two ways for interference to arise. First, the eye must accommodate to see clearly with each change in focal distance that occurs with each eye movement to a new fixation along the curved surface. Since accommodation changes are relatively slow, speed of reading may be retarded when reading is done under such conditions. Secondly, when part of a curved line of print is farther from the eye than other parts, the letter and word forms along the line are distorted. Although this distortion does not become important when the printed page is nearly flat, it may seriously interfere with visibility and hence with clear and rapid perception when there is considerable curvature of the printed page.

The specific problem in this study is to investigate changes that occur in speed of reading and in visibility of words when the reading copy is slanted away from the eyes at different angles. The results should provide some evidence concerning the effects of accommodation changes and distortions of the visual images of words upon ease and speed of perception in reading curved lines of print in large books. Speed of perception will be measured in terms of speed of reading. The effect from distortions of word forms on ease of perceiving will be measured in terms of visibility of selected words.

¹ This research was supported by a grant from the Graduate School, University of Minnesota.

MATERIALS AND PROCEDURE

The reading material consisted of Forms I and II of Tinker's *Speed of Reading Test*². In each form there are four hundred and fifty items of thirty words each. As a check on comprehension, the reader crosses out the one word that spoils the meaning of each item. Since responses are more than ninety-nine per cent correct, the test virtually measures speed of reading with comprehension constant. The two forms, which are approximately equivalent, were printed in ten-point Excelsior type in a twenty pica line width with two-point leading on eggshell paper stock.

For the visibility measurements, thirty five-letter words were cut from the test and pasted on white cards, one word per card. All words used are within the five hundred most familiar words in the Thorndike and Lorge list³ and consequently have a high index of familiarity.

The experiment was conducted in a laboratory which provided twenty-five foot-candles of indirect, well-diffused incandescent illumination. The subjects were one hundred and eighty university sophomores. They were divided into three groups; one control and two experimental groups. The subjects were tested individually, and the order of testing was systematically varied.

A reading stand with the reading surface at forty-five degrees to the table was employed for holding the flat copy that was to be read in the normal manner (Test Group 1, Control). Two triangular copy holders were devised to hold the reading text with the lines at a slant; part of the line near subject, part farther away. The triangular copy holder was placed upon the regular reading stand and so arranged that the center of the slanted lines to be read were the same distance from the eyes as the copy read upon the regular reading stand. One triangular copy holder slanted away from the flat reading plane (and consequently from the eyes) at forty-five degrees (Experimental Group 2) and the other at sixty degrees (Experimental Group 3).

On arrival at the laboratory a subject was adapted for five minutes to the illumination employed while taking the practice exer-

² Not yet available for general distribution.

³ E. L. Thorndike and I. Lorge. *The Teacher's Work Book of 30,000 Words*. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

cise of the reading test. In Test Group 1, the control group, subjects read Form I and Form II of the reading test on the flat surface of the regular reading stand, i.e., normal reading situation. In Experimental Groups (2 and 3), Form I of the test was read in the same position as with the control group. For Group 2, Form II of the test was read slanted at forty-five degrees, and for Group 3, Form II was read slanted at sixty degrees. The time limit was five minutes on each test form in all testing. The subject responded by speaking aloud the word that spoiled the meaning in each item.

After the speed of reading tests were finished, visibility measurements were made with the Luckiesh-Moss Visibility Meter. These measurements were made under the same conditions as the speed of reading measurements, i.e., on a flat surface, slanted at forty-five, and at sixty degrees. After six practice words, the visibility of eight words was measured in each position. The order of presentation was systematically varied.

RESULTS AND DISCUSSION

The basic data for the speed of reading measurements are given in Table 1. The mean scores in column 4 represent the number of items of thirty words each read in five-minute periods. In the control group there was a statistically insignificant gain of .60 item on Form II. The means in the experimental groups were 'corrected' by this amount. The reliabilities of the scores (Form I vs. Form II) are given in column 8. They range from .92 to .96.

In Group 2 are the results for reading flat copy in comparison with text at forty-five degrees from the flat surface. Material at the forty-five degree angle was read 5.7 per cent slower. The critical ratio in column 9 reveals this to be a highly significant difference. In Group 3 are the results for reading flat copy vs. copy at a sixty degree slant. Here the slanted copy was read 16.4 per cent slower. The difference is highly significant.

Data for the visibility measurements are given in Table 2. The scores are in terms of scale readings on the visibility meter. The consistency of readings from one condition to another was high. The coefficients in column 6 range from .81 to .93. The split-half reliabilities for responses within each of the three conditions were also computed. These (uncorrected) are: flat copy, $r = .93$; forty-five degree slant, $r = .88$; sixty degree slant, $r = .91$.

Examination of the data in Table 2 shows that the visibility of

words is reduced markedly when the words are at an angle in comparison with flat copy. Thus in column 5, there is a 31.3 per cent loss when the slant is forty-five degrees, and a 48.4 per cent loss when the slant is sixty degrees. The sixty-degree slant is 24.9 per cent worse than the forty-five-degree in reducing visibility. All these differences are highly significant as shown by the corresponding critical ratios in column 7.

The results in Tables 1 and 2 show that readability of print is

TABLE 1.—SPEED OF READING TEXT ON FLAT AND ON SLANTED SURFACES

Group	Test Form	Position of Copy	Mean	SE _M	Corrected Difference*		r	D/SE _{Diff.}
					Paragraphs	Per Cent		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	I	Flat	52.50	1.42	0.00	0.0	.96	—
	II	Flat	53.10	1.37				
2	I	Flat	52.95	1.58	-3.02	5.7	.95	5.93
	II	45° Slant	50.53	1.58				
3	I	Flat	52.17	1.65	-8.54	16.4	.92	12.77
	II	60° Slant	44.23	1.37				

* The differences in column 6 are 'corrected' by the amount of the difference between the mean scores of Form I and Form II of Test Group 1 which serves as a control group. The 'correction' amounts to 0.60 items for each test group comparison and is subtracted from the Form II score.

TABLE 2.—VISIBILITY MEASUREMENTS OF WORDS ON FLAT AND ON SLANTED SURFACES

Comparison	Mean	SE _M	Diff.	Per Cent Difference	r	D/SE _{Diff.}
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Flat vs. 45° Slant	4.44	.25	1.39	-31.3	.91	9.93
	3.05	.13				
Flat vs. 60° Slant	4.44	.25	2.15	-48.4	.81	12.13
	2.29	.10				
45° Slant vs. 60° Slant	3.05	.13	0.76	-24.9	.93	14.11
	2.29	.10				

adversely affected to a marked degree when the copy is held so that parts of the lines of print are at different distances from the eyes. This is true for both speed of reading and for visibility of the words forms. Although accommodation changes occur during the slanted reading, it is difficult to estimate the relative importance of these changes in this experiment. The effects of the slanted print on visibility are so great that they would seem to be dominant in reducing the speed of reading. Apparently, a modified design will be necessary to get at the accommodation effects.

A closer approximation to the practical situation would be to study reading on a curved surface vs. a flat surface. This study is now under way.

The results obtained in this study suggest that the marked curvature of lines of print near the inner margin of large books and journals can adversely affect readability. If much wider inner margins were used, the situation would be much improved.

SUMMARY

1) This experiment was designed to investigate the effects upon speed of reading and upon visibility of word forms when the text is slanted so that parts of a line are at varying distances from the eyes.

2) Material slanted at forty-five and at sixty degrees was compared with flat copy. There were one hundred and eighty subjects.

3) Speed of reading was retarded 5.7 per cent by the forty-five-degree slant, and 16.4 per cent by the sixty-degree slant.

4) Visibility of word forms was reduced 31.3 per cent by the forty-five-degree slant, and 48.4 per cent by the sixty-degree slant.

5) The rôle of accommodation changes cannot be determined from these results.

6) It is suggested that a major portion of the retardation in reading speed is due to the reduced visibility of the words.

7) These results have a bearing upon printing and binding of large books and bound journals where there tends to be considerable curvature of the lines of print near the inner margin. To avoid the deleterious effects of this curvature upon readability, it is suggested that large books and journals should be printed with a much wider inner margin.

DELINEATING THE AREAS OF EDUCATIONAL AND CLINICAL PSYCHOLOGY

RUSSELL N. CASSEL

Major, USAF, Scott Air Force Base, Illinois

The training and duties of the educational and clinical psychologists are two separately organized and distinct disciplines, but they are very closely related, and there is considerable overlapping of their respective areas. The primary objective of the educational psychologist¹ is to maintain the mental good health of individuals through the psychoeducational or mental hygiene clinic and community mental hygiene programs (2, 15); while the clinical psychologist deals primarily with mentally-ill persons emphasizing therapy procedures and related practices through the psychiatric or psychological clinic. The educational psychologist deals primarily with teachers and religious ministers through the public schools, churches, and other social and industrial organizations; while the clinical psychologist deals with the psychiatrist, medical doctor, and other members of the psychiatric team through the hospitals, mental institutions, penal institutions, and other agencies charged with the care and rehabilitation of mentally-ill persons (1, 5, 14).

As the increased complexity of the continuous specialized development of these two psychological disciplines emerges, the need for a more precise delineation of their respective areas of activity becomes more evident. This article defines the boundaries of these two important disciplines based on functional objectives as they are visualized by the author (5, 6, 7).

'Happiness' of the human being, in its broadest sense, is the ultimate goal for both of these disciplines, and good mental health is a necessary prerequisite for this happiness. The psychological structure for human happiness consists of an on-going process of goal-setting and goal-striving in which ego-involvement is present (7, 13). In addition, optimum happiness of the individual must insure the greatest space for freedom of movement of the individual

¹ The term 'educational psychologist' as it is used in this paper refers to the practice rather than the teaching; the practitioner rather than the professor.

possible predicated on a minimum of infringement of rights of others. It thrives best in a social climate effervescing with self- and socially-imposed barriers to which all individuals are vulnerable (7).

As the educational and clinical psychologists maintain or regain and restore the mental good health of individuals, they both work primarily with the personality. The first working primarily with typical individuals, and the latter with atypical and abnormal ones. Impinging psychological dynamics of a traumatic kind seriously impair the mental health of the individual; the after-effect of these emotional experiences leave mental and emotional scars that impair the mental health of the individual equally as much as the loss of a leg or an arm would impair the physical functioning of the organism. Just as the prepared vaccines of preventive medicine reduce the rate and consequence of physical illness so, effective mental hygiene programs may prevent and control the spread of mental illnesses.

Statistical data indicate that the precipitation of mental illnesses increase rapidly after the age of sixty, and even greater after seventy. In 1900 only a small percentage of our population lived beyond the age of sixty—about four per cent; today that figure has increased to more than seven per cent, and it is anticipated that in the next ten years more than 11 per cent of the population will live beyond age 60. It may be that the general deterioration of organic functioning with age makes the individual more vulnerable to mental illness; or it is possible that scientific progress in the prevention and treatment of mental illness has not progressed as rapidly as for physical illnesses (2).

In 1946 by passing the Federal Mental Health Act, the United States has accepted a large portion of the responsibility for the maintenance of good mental health for its citizens just as it had previously with venereal disease, tuberculosis, small-pox, cancer, and other physical diseases. The goal of this act is to insure every American citizen the opportunity for good mental health by providing federal support for the development of mental hygiene preventive treatment programs in our local communities. Increased emphasis has been placed on preventive procedures that will insure good mental health for all individuals, and to provide the necessary training and research to implement this program. The increased use of the educational psychologist in the preventive aspects of this

program, has resulted in an immediate need to delineate between the areas of these two disciplines.

I. THE EDUCATIONAL PSYCHOLOGIST IN THE PSYCHO-EDUCATIONAL CLINIC

Educational psychology as a science stems from the work of Pestalozzi, Herbart, and Froebel in their attempts to psychologize education by revising the methods for the presentation of subject matter during the late 19th Century. The second major movement in its development was the organization of a course for teachers and prospective teachers in educational psychology by E. L. Thorndike, Charles Judd, and Lewis Terman about the same time and around the turn of the 20th Century. Since about 1900, educational psychology has been concerned with three major areas: (1) research and the analysis of educational problems, (2) the development and organization of principles, approaches, and theories of learning, and (3) the guidance and counseling of individuals educationally, vocationally, and personally. From this last area, and stemming directly from the impetus and need furnished by the Mental Health Act, has emerged the concept of educational psychology as described herein.

The Psycho-educational or Mental Hygiene Clinic.—In accordance with the provisions of the Mental Health Act, the mental hygiene clinics sponsored by federal money emphasize preventive measures designed to maintain the mental good health of all individuals. An outline of the principal functions imperative to such an effective preventive program, and in which the educational psychologist becomes the principal figure, is as follows:

I. Educational and Vocational Guidance

A) Educational Planning and Guidance:

- 1) Junior high school level.
- 2) Senior high school level.
- 3) College level.
- 4) Graduate school level.
- 5) Adults.

B) Vocational and Occupational Planning and Guidance:

- 1) Integrating the educational and vocational plans.
- 2) Integrating the 'life-plans' of the individual.

C) Life-planning for late adolescent and post adolescent individuals.

II. Counseling, Psychodiagnoses, and Therapy for the Typical Personality Class of Individuals

A) Psychodiagnoses and Prognoses of Mental Illness, and the Referral of Severe Clinical Pictures.

B) Individual and Group Psychotherapy.

C) Authority, Parental, and Personal Aid Relative to Child Delinquency and Discipline Problems.

D) Aid in Dealing With Inter-personal Conflict Problems:

1) Child-parent conflict.

2) Employer-employee conflict.

3) Teacher-pupil conflict.

4) Peer conflict.

5) Sibling conflict.

6) Marital conflict, and sexual adjustment.

III. Mental Hygiene Programs for the Individual or the Community for Maintaining the Mental Good Health of the Typical Personality Class

A) Pre-marriage Planning:

1) Prospective bride.

2) Prospective groom.

3) Prospective 'in-laws' and parents of bride and groom.

B) Pre-family Planning and Birth Control.

C) Pre-natal Parental and Family Adjustment Program.

D) Infant Care and Child Development.

E) Child Guidance, Discipline, and Mental Hygiene Care.

F) The Puberty Cycle and Adolescent Development.

G) Adult Life Planning Programs.

H) Marriage and Sex Adjustment for Adults.

I) Middle-age or Menopause Adjustment Program.

J) Old Age or Senile Adjustment Program.

K) Sex Education and Orientation for the Child:

1) The six-year old.

2) The sixth-grader.

3) Post-puberty cycle.

L) Aiding Individuals to Reach Higher Levels of Adjustment.

IV. Professional and Consultant Services

- A) Conduct Research Related to the Prevention of Mental Illness and the Maintenance of Mental Good Health.
- B) Conduct Research Related to Learning and the Learning Process.
- C) Professional Surveys on Matters Pertaining to Learning.
- D) Lectures and Consultant Services Relative to Educational Psychology.
- E) Organization and Supervision of the Training for the Assistants and Interns in the Field of Educational Psychology.

II. THE CLINICAL PSYCHOLOGIST IN THE PSYCHIATRIC CLINIC

The discipline of clinical psychology has emerged emulating and keeping parallel with the development of clinical psychiatry during the early portion of World War II. In the most progressive clinics, the trend is to function in conjunction with other members of the psychiatric team. The team is usually made up of the following members: (1) neuro- or clinical psychiatrist, (2) a clinical psychologist, and (3) a social case worker. In general, the psychiatrist heads the team and is responsible for coördinating and integrating the services of the other members, and for determining the diagnoses, prognoses, and final disposition of the case. The clinical psychologist performs a function for the psychiatrist similar to that of the medical laboratory for the internal medical practitioner in that he administers and interprets psychological tests which are designed to reveal the presence of incipient and latent conditions of psychoses, and assists the psychiatrist in other psychological procedures designed for the treatment and prevention of mental illness. The social case worker is primarily concerned with working-up and writing the case history of the individual, but, also, assists the psychiatrist in psychological procedures designed for the treatment and prevention of mental illness.

Motivated by a desire to retain their patients, many medical doctors have established the practice of referring mentally ill persons to a psychological clinic headed by a clinical psychologist. Under this set-up, the clinical psychologist assumes full responsibility for all psychological aspects, and continues to work with the patient in conjunction with the medical doctor making the referral,

who retains the responsibility for the physical aspects of the care and treatment.

The Psychiatric or Psychological Clinic.—The clinical psychologist deals primarily with those individuals who are diagnosed as being mentally ill, and the emphasis is placed on meliorative or curative measures which are designed to regain the good mental health. Because of the complex nature of mental health and its interrelatedness with the physical health of the individual, a basic fundamental in the meliorative approach is to investigate the presence of physical aspects either before or concomitant with the psychological approach. In line with this kind of policy, the following areas of concern constitute the primary functions of the clinical psychologist:

I) Psychodiagnostic Collaboration with the Psychiatric Team or a Medical Doctor for Mentally Ill Persons from the Atypical and Abnormal Personality Classes:

- A) Problems of incipient or latent psychoses.
- B) Problems of intra-cranial organic pathology.
- C) Severe clinical pictures of psychoneuroses.
- D) Deterioration or impairment of mental functioning.
- E) Problems of sexual aberrations.
- F) Penal and severe delinquent problem cases.
- G) Inter-personal dynamic problems of a severe type.

II) Meliorative Collaboration with the Psychiatric Team or a Medical Doctor for Individuals from the Atypical or Abnormal Personality Classes:

- A) Individual or group therapy.
- B) Pre- and post-neurosurgery evaluation and planning.
- C) Rehabilitation and reëducation of neurosurgery patients.
- D) Psychoanalyses (where the clinician is properly qualified).
- E) Penal and delinquency rehabilitation.
- F) Plan, organize, and direct an occupational therapy program.

III) Organization and Direction of Speech Clinic:

- A) Psychodiagnostic collaboration and treatment of aphasic, apraxia, agnosia, and other related problem cases.
- B) Collaboration in the diagnoses and treatment of other problems of speech disfunctioning and defects.

- C) Collaboration in the planning and correction of speech defects.
- IV) Professional and Consultant Services:
 - A) Conducts research related to the diagnoses and treatment of mental illness, and its prevention.
 - B) Lectures and consultant services relative to clinical psychology.
 - C. Organization and supervision of the training for assistants and interns in the field of clinical psychology.

III. TRAINING

Just as the child of today becomes the parent of tomorrow; so the typical individual of one day becomes the atypical and abnormal individual of another day. The training of the educational psychologist emphasizes preventive and prophylactic procedures and functions; while that for the clinical psychologist, emphasizes curative or meliorative ones. There is certainly no province or aspect in the area of psychological procedures and techniques that is solely the concern of either of these separate disciplines. In the present program of training, as indicated by the present need, there appears to be a definite trend in the training of educational psychologists to include more and more data relative to projective techniques and their use in the clinic, and the provision for clinical internships. For clinical psychologists, the training tends to include more and more data relative to the psychology of learning, and the rôle of learning in effective mental hygiene preventive programs.

SUMMARY

This article is concerned with delineating the areas of the educational and clinical psychologist as two separate disciplines, and describes the functions of each as they are visualized and practiced by the author. The Federal Mental Health Act of 1946 has tended to emphasize the rôle of the educational psychologist in the prevention of mental illness among individuals. Thus, the rôle of the educational psychologist tends to center around the preventive or prophylactic function relative to mental illness; while the clinical psychologist is concerned primarily with curative or meliorative ones.

REFERENCES

- 1) *Air Force Manual 160-45*, Military Clinical Psychology. Departments of the Army and the Air Force, July 1951.
- 2) H. W. Bernard. *Mental Hygiene for Classroom Teachers*. New York: McGraw-Hill Book Co., 1952.
- 3) M. F. Berry and J. Eisenson. *The Defective in Speech*. New York: Crofts.
- 4) N. Blackman. "Group psychotherapy with aphasics," *Journal of Nervous and Mental Disease*, Feb., 1952, pp. 154-63.
- 5) R. N. Cassell. *An Experimental Investigation of the 'Reality-Strata' of Certain Objectively Defined Groups of Individuals by Use of the Level of Aspiration Technique*. Los Angeles, California: Ed.D. Dissertation, University of Southern California, 1949.
- 6) ———. *Manual—The Cassel Group Level of Aspiration Test*. Beverly Hills, California: Western Psychological Services, 1952.
- 7) ———. *Manual—The Cassel Psychotherapy Progress Record*. Beverly Hills, California: Western Psychological Services, 1953.
- 8) R. M. Dorcus and G. W. Shaffer. *Textbook of Abnormal Psychology*. Baltimore, Maryland: Williams & Wilkins Co., 1945.
- 9) A. I. Gates and others. *Educational Psychology*. New York: The Macmillan Co., 1949.
- 10) E. L. Hartley, H. G. Birch, and Ruth E. Hartley, *Outside Readings in Psychology*. New York: Thomas Y. Crowell Co., 1950.
- 11) D. O. Hebb. *Organization of Behavior*. New York: Wiley, 1949.
- 12) J. McV. Hunt. *Personality and the Behavior Disorders*. New York: The Ronald Press Co., 1944.
- 13) K. Lewin. *A Dynamic Theory of Personality*. New York: McGraw-Hill Book Co., 1935.
- 14) L. A. Pennington and I. A. Berg. *Introduction to Clinical Psychology*. New York: Ronald Press. 1948.
- 15) E. L. Shostrom and L. M. Brammer. *The Dynamics of the Counseling Process*. New York: McGraw-Hill Book Co., 1952.
- 16) S. R. Slavson. *The Practice of Group Therapy*. New York: International University Press, 1947.

A COMPARISON OF THE VALUE SYSTEMS OF SOUTHERN NEGRO AND NORTHERN WHITE YOUTH

JULIUS H. HUGHES and GEORGE G. THOMPSON

Syracuse University

The purpose of this study was to compare the values, as measured by value-analysis, of a group of Negro boys and girls living in the South and attending a racially segregated school with those of an equal number of white adolescents living in the North and attending a non-segregated school.

EXPERIMENTAL PROCEDURE

The technique used in the present study was introduced by White who made a value-analysis of a Negro novelist's autobiography (1). Value-analysis as developed by White consists of three steps:

- 1) Put in the margin a symbol corresponding to each goal and each value-judgment that is explicitly stated in the material, or clearly implied by it.

- 2) Tabulate these symbols.

- 3) Interpret each numerical result in the light of the picture as a whole, with special attention to the person's possible reasons for conscious concealment or unconscious self-deception (2).

The values which were thus analyzed are listed in Table I. The list was derived from White's fifty basic categories of value-analysis (2).

The writings that were analyzed for their value judgments were themes the experimental subjects were asked to write. The title given them was "Things I Would Like to Change in My Way of Living." These essays were written in class with instructions to the teachers to limit the writing time to exactly thirty minutes. The teachers were asked to make no further interpretation of the subject on which the students were requested to write. The boys and girls were free to define the subject in any manner that they wished.

It was anticipated that the students might avoid using certain words—which might prove to be key ones—because they were in doubt about the spelling of the words. The teachers were, therefore,

asked to point out to the pupils that they should spell a word the best they could whenever they were in doubt about its spelling.

Each student recorded his name, age, sex, father's occupation, mother's occupation, and length of time he had lived in the community. The time for recording this information was not included in the time allowed for the writing of the themes. No student who had resided in his present city for less than one year was included in the study. The teachers in the non-segregated school system identified for the investigators those essays which had been written by white students.

Twelve hundred and sixty-nine themes were obtained. These were grouped according to locale, sex, race, and age. Since there were fewer Negro boys in the non-segregated schools than in any other group, the data were examined to determine what adjacent chronological ages had the largest number of cases. These were found to be the fifteen- and sixteen-year-olds. There were twenty-two subjects in the smallest race-locale-sex subgroup of this age classification. All of the remaining essays written by fifteen- and sixteen-year-olds were grouped according to race, sex, and city, and each of these themes was assigned a number. Using a table of random numbers, twenty-two themes were selected from each group. This gave a total of eighty-eight essays. These were analyzed by the procedures recommended by White in his version of value-analysis.

In order to determine scorer reliability, nine of the themes which were not selected for the final analysis were analyzed independently by a third-year graduate student in psychology. These independent judgments were compared with those of the analyst who scored all of the eighty-eight themes in the experiment proper. A disagreement was recorded when one analyst recorded a symbol and the other did not and when different symbols meaning different values were used. The following sentence taken from one of the themes marked with symbols employed in the study provides an example:

Analyst 'A'

('Se) Everyone today is living in constant fear of a

('S) Third World War which probably will kill nine-tenths of the people in the world.

Analyst 'B'

('Se)

The independent judges agreed that the proper interpretation of this response is "there is no security." Analyst 'A', however, saw

an additional value of 'safety' and indicated that the statement suggested 'no safety'. These were recorded, therefore, as one agreement and one disagreement. The obtained percentages of agreement for the nine themes were as follows: 76, 81, 67, 69, 66, 78, 80, 65, 66. Approximately three months after the initial scoring of these nine themes, the principal analyst scored them again independently of his first scoring. The obtained percentages of agreement were as follows: 68, 91, 78, 76, 83, 84, 88, 89, 86. These are in keeping with the percentages of agreement reported by White.

RESULTS

Comparisons were made between the Negro boys and girls of the southern community and the northern white adolescents in terms of the frequency of the various values reflected in their themes. A cursory inspection of Table I reveals that there are a number of values which received a similar amount of emphasis from the two groups. Both groups gave substantial attention to knowledge: 8.2 per cent of the total emphasis units for the whites as compared with 8.0 per cent for the Negroes. The white students demonstrated an interest in money, high wages, and low prices as is evidenced by 5.6 per cent units on the economic value. So did the Negroes, 5.6 per cent emphasis units. A common interest in success and superiority is suggested by 6.6 per cent of the total emphasis units in the white youth's responses reflecting achievement, and 7.4 per cent emphasis units of the same type in the Negro youth's responses. That both the Negro and the white subjects were interested in having friends and being liked is suggested by the fact that they gave 3.4 per cent and 4.6 per cent, respectively, of their attention to the friendship value.

The data were tested for statistical significance of the differences in the proportion of emphasis given the values by the two racial groups. Where the theoretical frequencies were less than five, probability was determined by the direct method.¹

The Negro group gave considerably more emphasis to justice than did the white boys and girls ($p < .0001$). This implies that the Negro adolescents were far more concerned with equality of oppor-

¹ The formula used to determine these probabilities was:

$$p = \left(\frac{(a+b)!(a+c)!(b+d)!(c+d)!}{N!} \right) \left(\frac{1}{a!b!c!d!} \right)$$

TABLE I.—EMPHASIS ON DIFFERENT VALUES
(Northern Whites from Non-Segregated Community; Southern Negroes from Segregated Community)¹

Values	Percentage of All Emphasis—Units		Values	Percentage of All Emphasis—Units	
	White	Negro		White	Negro
Knowledge	8.2	8.0	Adjustment	1.0	1.2
Happiness	8.0	2.6*	Aggression	1.0	2.0
Security	7.6	3.8*	Beauty	1.0	0.8
Family-love	7.2	1.8*	Recognition	1.0	1.8
Achievement	6.6	7.4	Rest	1.0	0.2
Economic value	5.6	5.6	Self-regard	1.0	0.6
Independence	5.6	2.4*	Cleanliness	0.8	1.6
Safety	5.0	2.4	Food	0.8	0.6
Friendship	4.6	3.4	Pleasant personality	0.8	0.2
New experience	4.4	3.2	Appearance	0.6	0.8
Activity	3.4	2.0	Group Unity	0.6	5.0*
Work	2.8	4.8	Culture	0.4	0.0
Justice	2.6	11.8*	Dominance	0.4	0.2
Religion	2.4	3.4	Sex-love	0.4	1.8
Tolerance	2.4	2.0	Manners	0.4	0.2
Generosity	2.2	2.2	Likeness	0.2	0.6
Morality	2.0	3.8	Obedience	0.2	0.6
Comfort	1.4	1.8	Truthfulness	0.2	0.6
Health	1.4	1.0	Creative self-express-	0.0	0.2
Intelligence	1.2	2.4	sion		
Ownership	1.2	2.4	Excitement	0.0	1.4
Practicality	1.2	0.2	Sex	0.0	0.2
Value in general	1.2	1.2			

¹ N is five hundred values for the white boys and girls and four hundred and ninety-nine for the Negro youth.

* Statistically significant difference with a p value of at least .01. The probability of obtaining six or more significant differences by chance alone is .001.

tunity and fairness than were the white subjects. The attention of the Negro subjects was directed to the group unity value with a greater frequency than was that of the white group ($p < .0001$).

The white subjects were significantly more concerned with being content and happy than were the Negro students. This is indicated by the difference in the emphasis given to the happiness value ($p = .0002$). Closely related to this difference is the amount of emphasis that the two groups gave the security value. The white

boys and girls were more concerned with freedom from worry than was the Negro group ($p = .0098$).

The kind of companionship that is given by one's family (family-love) was given a significantly larger percentage of emphasis by the white than by the Negro adolescents ($p < .0001$). At the same time, however, the white subjects mentioned the *independence* value more than did the Negroes ($p = .0098$).²

The above analysis is in terms of the frequency with which the two groups mentioned the values. Probably of more significance is how the subjects treated the values. Value-analysis provides a means of dividing the values into four categories: frustration and disapproval, satisfaction, desires, and means-end relationships.

The data were examined for possible differences between the two racial groups in the satisfaction category. A significantly larger percentage ($p = .0044$) of the white subjects indicated satisfaction in terms of friendships. The white youth also enjoyed greater happiness ($p < .0001$) and family-love ($p = .0098$). The probability of obtaining three or more significant differences by chance alone is .001. The following statements are illustrative of how the white boys and girls treated the above values:

I am quite satisfied with my way of living. I have a very nice home, very good parents a good school in which to go and teachers who are swell.

Our family is an average family, we enjoy living and are very happy.

I am known to be very cheerful and patient among my friends.

For no value did a significantly larger percentage of the Negro pupils experience satisfaction.

An analysis of the frustration area reveals that a significantly larger percentage of the Negro boys' and girls' emphasis-units dealt

² The possibility was considered that the greater emphasis on the values by one group was the result of a few individuals each mentioning a particular value many times. The number of persons mentioning the six values reported as significant was compared, and the following p values were obtained: happiness, $< .0008$; security, $< .0110$; family-love, $< .0003$; independence, $< .1292$; justice, $< .0010$; group-unity, $< .0005$. Thus it can be seen that the differences between groups are not dependent on multiple responses by a few individuals.

with a lack of justice ($p < .0001$). For instance:

There are many other things I could change but I think this one is more important. Our school system, I think we are just as good as the next person. There should be no segregation in our school, and so it is we should have the same rights.

Always, every where we go whites are always first.

I would like to change the way colored people have to ride in the back of the buses.

The white adolescents appeared to be more frustrated in the areas of knowledge ($p = .0070$) and achievement ($p = .0038$). For example:

I have failed algebra and dropped Latin. I am afraid I will not be able to get into college even tho I am going to take 2 years of Spanish.

Longer vacations on holidays. Less school books published.

I would change my attitude toward studying. I would get down to some hard studying instead of watching television all night long until it is to late to study for the test the next day.

In addition, the white group emphasized a lack of security and inadequate safety more than did the Negro boys and girls. For example:

Wars are the largest ways in which nations have tried to gain what they want. Ever since man came on earth there have been wars in which a few have profited and the majority have lost.

There is one very important change and that is, that I would like to have grown up in a world which had not known war.

Some are afraid that their sons and daughters and husbands are going to die.

They (old people) are put in homes, and other institutions where they have nothing to do but sit and walk around unwanted, unloved, and useless.

The probability of obtaining five or more significant differences by chance alone is .001.

When the two areas, frustration and satisfaction, are considered together, it appears that the southern Negro youth find the environment more conflict-producing than do the northern white boys and girls. This is suggested by a frustration-satisfaction ratio of

3.00 for the Negro pupils as compared with one of 1.19 for the white students. It seems that the Negro adolescents perceive segregation as a condition which produces inequalities, for most of their references to a lack of justice contended that it is 'unfair' to have to sit in a certain section on a bus, to have separate schools, etc. Justice, moreover, accounted for approximately one-third of the Negro students' emphasis-units in the frustration area.

Of the forty-one values mentioned in the desires of the two groups, there was a statistically significant difference (at $p < .02$) for four of the desires. The probability of obtaining four or more significant differences by chance alone is .008.

Both groups emphasized the desire for knowledge more than they did any other value. The Negroes tended to perceive knowledge as a means of self-enhancement and to use that value to attack 'unequal educational opportunities'. The following statements were taken from the themes written by Negro students:

I'd like to have it said when I graduate from this school that, that is one of our best.

...if there should come a time that we can go to school with the other races we should be given the same rights and there should be some Negro teachers, and with the same privilege.

I would study a little harder than I have so I won't have to cram everything in my head at one time and stay up half the night trying to get it.

Another thing I would like to change and that is equal right of an education.

I would endeavor to study more. There are some students I idolize for their brilliant work in school.

I plan to continue my education beyond high school and even beyond my B.S. degree.

The white boys and girls, too, appeared to be interested in knowledge as a means of extending the ego. There is some indication that they also feel that the school is not just. For example:

First I think that the periods should be only a half hour long and after each period we should be allowed an half hour study hall in which to do our homework.

I would like to get higher marks and be a better student, most of all I would like to get a scholarship for college—if it were possible.

I try to get fairly descent marks but I just can't seem to do it. I study every night for more than an hour at least.

I would like to go to college but I am not smart enough.

Both groups gave substantial emphasis to the economic value (a rank of two for the white pupils; four, for the Negro pupils).

The Negroes gave significantly more emphasis to the achievement value. Moreover, some of them appeared to view achievement as at least a partial solution to their problems. Examples follow:

I think boys who are going to high school or college should not be drafted. They should be allowed to finish. Because the army force want men who are educated.

I would perform all my duties to the best of my ability, and I would study so that I would be a better student than which I am.

Oh, if I could ever learn to type.

I would first of all drop all my subjects except cosmetology and complete that course. After finishing cosmetology and getting my licence I would continue and finish school.

Last, I will strive and study hard to finish high school and try to enter college in the fall of '52, and to repay my parents for all they have done for me, succeed in my ambition to be a nurse in the United States Armed Forces.

The Negro group also gave more emphasis to the following values: justice, work, and group unity. These southern Negro adolescents appeared to see a substantial relationship between these values, for they tended to treat each in terms of providing equality of opportunity. Some examples:

I would get a few more theaters, and I would also have mixed schools oh well mixed everything because it really doesn't matter what color you are it is what one is on the inside so I would do away with this racial problem.

I, as an American, would like to be treated as an American. In the Constitution there is a law which says "All men are created equal". If this is true please I ask to God to let it be fulfilled.

I would like for Negroes to have the same chances and opportunities that people of the white race have. Such as: If I trained for the same position as a white person and we both

applied for the job and were qualified, just because I am a Negro and he is white, the white person would get the job.

I would like for the relationship between the Negroes and whites to change. I would like to get a job on week end days like the white boys. Maybe if I had a chance to mix with the white boys and girls now it would be a great help.

I would like to see the time when all men united together and help make the United States be a larger and better place to live.

Negroes Helpt to build the United States and look at the privologes they have they must go to thear own theathers yet the foreignor goes to the white and colord shows.

Another change would be, is getting better jobs for young and old Negro people.

The data were not analyzed for differences in means-end relationships because both groups tended to view the values as ends in themselves with little expressions of specific means to reach desired goals. This may have been a function of the topic on which the boys and girls were asked to write.

CONCLUSIONS AND DISCUSSION

This analysis of the values of the white adolescents living in the North and attending a non-segregated school and those of a group of Negro boys and girls living in the South and attending a racially segregated school reveals that there is a tendency for both groups to place a similar amount of emphasis on the great majority of the values. The most outstanding differences in their values are in the following areas: justice, family-love, happiness, and group unity.

The emphasis upon group unity suggests that the southern Negro adolescents have a strong drive to be accepted as an essential part of the community. The Negro group tended to identify strongly with their own ethnic group and appeared to be quite egotistic. Most of their references to injustice, for example, continually demanded 'equal rights' for Negroes. Many of the statements, in fact, implied that the Negro is the only group that is not being treated fairly. Identification with the Negro race may have been in this instance a convenient and socially acceptable means of expressing a desire to secure equal rights for the individual. These southern Negro boys and girls were straight-forward in their criticism of segregation. This tends to invalidate the hypothesis held by some casual

observers that the southern Negro is content and happy with his present status. The theme of social injustice recurred again and again in the racial-segregation setting. This finding is in accord with that of Wright (3) who concluded that "Negro college students largely disapprove of the traditional 'place' of the Negro."

The northern white boys and girls appeared to have a stronger identification with the family. These boys and girls were rather realistic in their treatment of this value. Whereas approximately fifty per cent of the emphasis units of this value indicated that they were enjoying family love, approximately twenty-five per cent suggested that family frictions did exist.

The northern white adolescents reflected their environment as being quite threatening. They did not, however, reflect it as threatening as did the Negroes. This obtained difference, which was distinctly favorable to the white youth on the happiness value, can probably not be interpreted in a simple straight-forward fashion, since there is some evidence that for the Negro youth the securing of social justice and group unity is synonymous with having happiness.

REFERENCES

- 1) R. White. "Black boy: a value analysis," *Journal of Abnormal and Social Psychology*, 1947, 42, 440-461.
- 2) ———. *Value Analysis: The Nature and Use of the Method*. New York: Society for Psychological Study of Social Issues, 1951.
- 3) H. Wright. *Racial Humor: A Value Analysis, Abstracts of Doctoral Dissertations*, The Ohio State University Press, 1947, No. 52, 331-338.

BOOK REVIEWS

JOHN W. M. WHITING AND IRVIN L. CHILD. *Child Training and Personality*. New Haven, Conn.: Yale University Press, 1953, pp. 353. \$5.00.

This extremely interesting and valuable work deals, as the title suggests, with some empirically determined relationships in a large number of primitive cultures among certain child-rearing practices and several adult behavior patterns conventionally subsumed under the term 'personality'. While cross-cultural studies have been used for many years to help understand and explain personality development, these investigations have rarely used psychological variables, such as habit or generalization, in their hypotheses. Few, if any of these previous investigations have attempted to subject their hypotheses to statistical tests with quantified data. Whiting and Child, in their use of conventional psychological variables and in their attempts at quantification, have clearly made two significant contributions to the usefulness of the cross-cultural approach for psychology.

The authors, a cultural anthropologist and an experimental psychologist, were interested in "the extent and ways in which personality processes in individual members of a society determine the integration of the culture." Their general hypotheses were couched, for the most part, in terms of reinforcement learning theory, using Hull's theoretical model. Neal Miller's seminal work on conflict and displacement plays an important rôle in several of the authors' theoretical analyses.

More specifically, the authors studied the child-rearing practices in seventy-five primitive societies on which relevant data were available in the ethnographic reports of the Human Relations Area Files. To quantify this child-rearing data, each culture was rated on the initial satisfaction and the socialization anxiety in five separate areas (or systems) of parent-child interaction—oral (i.e., weaning), anal (i.e., toilet training), sexual, dependency, and aggression. The ratings of initial satisfaction were based upon the length of initial gratification, the amount of freedom given to the child to perform his initial habits, and the amount of concurrent anxiety. The ratings of socialization anxiety or the concurrent anxiety involved in imposing inhibitory controls were based upon

the brevity of initial satisfaction, the severity and frequency of punishment, and by overt signs of emotional disturbance in the children. These quantitative ratings of initial satisfaction and socialization anxiety were then compared with the particular explanations of illness and the consequent therapeutic practices in the various cultures.

In drawing the specific hypotheses for testing much use was made of the concepts of fixation, identification, projection, and displacement. While drawing upon the original psychoanalytic formulations for a beginning, they then translated these notions to fit into their 'general behavior theory' much in the fashion of Dollard and Miller (*Personality and Psychotherapy*).

Of particular interest to psychologists is Whiting and Child's treatment of fixation. They believe that the consequences of extreme frustration and extreme gratification are quite dissimilar and that inclusion of both antecedent conditions under a general notion of fixation masks the potential usefulness of such an interpretation. They suggest that fixation resulting from a high degree of gratification be termed '*positive fixation*' while that resulting from a high degree of frustration be termed '*negative fixation*'. The concept of negative fixation is used in conceptualizing the relationship between socialization anxiety and explanation of illness, while positive fixation is used in conceptualizing the relationship between initial satisfaction and the therapeutic practices in illness. The notion of negative fixation would suggest, in this instance, that societies high in oral socialization anxiety should demonstrate this anxiety in utilizing oral explanations of illness (i.e., some act of ingestion on the part of the patient is responsible for the illness), while societies with low oral socialization anxiety should use such explanations less frequently. The concept of positive fixation would suggest that high initial satisfaction in an area should give rise to the performance of responses in that area as a therapy in illness, so that high initial oral satisfaction should lead to some sort of oral therapy. These two concepts of fixation are used in a similar fashion to study socialization anxiety and initial satisfaction in the other four areas of development (anal, sexual, dependency, and aggression) and the concomitant explanations of illness and therapeutic practices. The over-all empirical evidence is interpreted as justifying the separation between positive and negative fixation with strong evidence for a concept of negative fixation. The lack

of strong support for a concept of positive fixation is explained in terms of the possible inadequacy of use of "therapeutics as the criterion variable, the inadequacy of the data on therapeutic practices, and the more general problem of progressive satisfactions, i.e., the changing nature of fixated goals with maturity. The authors suggest that this dichotomy between positive and negative fixation be further studied in future investigations of this sort.

The text goes on to discuss the origins of guilt as a function of identification and the origins of the unrealistic fear of other persons and of spirits as a function of anxiety about aggression. Limitations of space prevent a fuller exposition of these hypotheses.

The authors come to two rather general but related conclusions in their final chapter. Firstly, the early life experiences of a person do have a persistent effect on this personality. Secondly, the personality characteristics of the typical members of a society do determine many of the beliefs and practices of that society. These conclusions are bolstered by the more specific findings partially discussed above.

The potential reader should be forewarned about one possible danger in reading this text. The highly sophisticated statistical procedures (e.g., normalizing and pooling several judges' ratings (p. 232 ff.)) might lull the unwary reader in forgetting the coarseness of the basic data and the real limitations such coarseness sets. While the authors mention this consideration, it remains a problem. The potential usefulness of this methodology coupled with the present positive findings suggest that any future investigations of this type would profit from the use of rather complete ethnological data collected with the specific variables to be studied in mind. In some places the authors' speculations and theorizing transcends the available data. For the most part, fortunately, these speculations are clearly labeled as such.

These minor criticisms in no way detract from the over-all excellence and importance of this text. It represents a significant contribution to behavior theory, child psychology and cultural anthropology. Specialists in these areas cannot afford to neglect this important contribution to our empirical knowledge and theoretical understanding.

LEONARD D. GOODSTEIN

State University of Iowa

G. WILSON SHAFFER AND RICHARD S. LAZARUS. *Fundamental Concepts in Clinical Psychology*. New York: McGraw-Hill Book Co., 1952, pp. 540.

Good textbooks in the field of clinical psychology are looked for with eagerness at the present time. Shaffer and Lazarus have tried with a great deal of success to provide an introduction to "basic principles, methodology, and general techniques." The jacket generously promises that the book "provides the training which is indispensable for understanding and competence in the specific clinical techniques." The authors, with whom this reviewer agrees, suggest more modestly that the book may "provide the foundation necessary for growth in proficiency and for more complete understanding of the administration and interpretation of the technical instruments of diagnosis and the methods of conducting a therapeutic interview."

The book very properly assumes that the reader will have as prerequisite, "some knowledge of general and abnormal psychology."

A brief historical survey brings one to the present situation in clinical psychology. The main subjects are dealt with in the following order: methodology; clinical techniques, intelligence and its measurement; defects of intelligence; the nature of personality; personality measurement; introduction to psychotherapy; psychotherapeutic devices; psychoanalysis; distributive analysis and synthesis; special psychotherapies; physical and chemical therapies; and the clinician in action.

One is perhaps impressed most with the excellent balance and good judgment found throughout the book. It is generally possible for the critic to point out over-emphasis on some subjects, and failure to properly include and discuss others; and although this is possible for the present volume, the more accurate criticism is that there has been excellent selection of material and fine judgment in its treatment.

If the reader is interested in the practical applications of psychology in one of its most important fields of application, he cannot afford to miss the critical summaries presented in this book; if he is interested in fundamental facts and concepts of clinical psychology, he may be surprised and pleased at the amount of general and historical information condensed in its pages. The condensed

form of writing does not mar the readability of the book but rather adds to its value.

It is interesting to note that although neither of the authors accepts the old 'faculty' psychology, (see p. 98), yet the old verbal habit can lead them to say that, "The reader is urged persistently to develop the critical faculty . . ." (p. vi). Which suggests that while the old determining tendencies (shall we say habits?) did not fail to show up in the writing, they might well have appeared more casually in the discussions. Shaffer or his co-author in another text wrote that, "The importance of habit formation in abnormal psychology is so great that it is not apt to be over-emphasized."

This most useful text not only presents to the reader well selected discussions of most of the fundamental matters which should be studied by the beginning student in clinical psychology, but the bibliography (pp. 495-525) puts him in touch with the remainder. It would have been of added value to the student if some of these could have been starred to indicate the most valuable collateral readings for the student. Pointing these out is perhaps the more appropriate task of the instructor.

Two well-made indexes complete the book. It will perhaps be the leading textbook for some years to come and is at the same time a valuable book of reference.

A. S. EDWARDS

The University of Georgia

OSCAR KRISEN BUROS, ED. *The Fourth Mental Measurements Yearbook*. Highland Park, N. J.: The Gryphon Press, 1953, pp. 1163, \$18.00.

Every testing specialist has learned to rely on the Buros Yearbooks as desk companions. Every teacher or counselor using published tests should study MMY as a whole, and consult it frequently. MMY is a bibliographer's dream: references to all tests, to all recent studies on each, and to books on testing. The organization and printing are magnificent.

The heart of the book is its critical reviews of tests. In the new Yearbook, these are invariably thoughtful. Except in the area of classroom achievement tests, reviewers have frequently tried out the tests and many reviewers report minor research performed to

gather new data for the review. Criteria for judgment are much the same as in former yearbooks. The reviews give less weight than formerly to picayune errors, more to the function of a test in instruction and clinical interaction. A long honor roll could be prepared of superior reviews. The following have exceptional general interest: a spectacular review by Shimberg of *Health Inventories* (ETS); Anastasi on 'primary abilities' tests; Westby on *Progressive Matrices*; Courtis on the *Thorndike Handwriting Scale* (some notable history); Fowler on *Henmon-Nelson*; and Kaulfers on *CEEB French Reading*.

The yearbook draws attention to new tests which might otherwise be overlooked. Judging from the reviews, the person interested in tests should become acquainted with these: 1947 *Progressive Matrices*, *California Cumulative Guidance Record* (Elem.), *Differential Abilities Test*, *SRA Self Scorer*, *IPAT Culture-Free Intelligence*, *GRE Aptitude*, *Diagnostic Reading Tests*, *Wing Musical Intelligence*, *Children's Apperception Test*, *Functional Evaluation in Mathematics*. While important, the *General Aptitude Tests of USES* and the *Weschler Scale for Children* were disappointing to reviewers.

There are a good many negative reviews.

Only three major publishers—World Book, Psychological Corporation, and ETS—are generally producing tests which the reviewers regard as adequately developed, and even these houses fail to maintain high standards consistently. Some other large publishers are accused of serious violations of judgment and even of ethics. Over and over, reviewers say that tests have been released prematurely, that they are acceptable for research use but because of insufficient validation or standardization are not suitable for practical use. This points up a major policy issue. Should we encourage the release of a test incorporating new ideas, even though the fundamental test development is incomplete? We need innovations, and yet we know that even the shiniest ideas may not survive careful research. Should research on a test precede general distribution, or follow it? Who should do it—the inventor or the customer? The reviewers demonstrate that we have gone too far in the direction of releasing tests prior to the required fundamental research. The majority of newer tests are at present suitable for use only by those who will conduct research on the tests in their own setting, but they are usually distributed without such limitation. A favor-

able sign is the report that many manuals have been revised to meet criticism from the last MMY. Older tests, however, have generally continued in print with neither revision nor continued research. Their authors would do well to withdraw some of them.

Buros has now perfected his yearbook technique. It is regrettably necessary to question the program and basic conception he holds, even though the series is marvellous if it can be sustained. The price of the volume is \$18, and this does not cover all costs. MMY is priced out of its market. If it cannot sell in volume, some new plan is needed. Changes can be made which will reduce the yearbook's value far less than they will reduce costs. An editor might consider these alterations. (1) Eliminate four hundred and fifty pages of reprinted book review excerpts, a costly feature not much used. (2) Eliminate reviews of foreign tests unlikely to be suitable for the American market. This is a provincial suggestion, but it would be better at this point to disseminate MMY more widely in this country than to pursue comprehensiveness. (3) Remove the exhaustive and expensive bibliographies on tests. (4) Eliminate analytic reviews of obsolete tests. A balanced review of (for example) the Zyve test is not needed. One sentence stating that this test should not be used for such and such purposes, signed if necessary by a group of adequate authorities, tells the student all he needs to know. Buros has this year made an effort to resurvey old tests, to such lengths as seven pages on variants of Army Alpha. Some retrospective reviews are of current value, but most could be reduced to much more blunt statements that a test is obsolete, or that, while originally 'experimental', in twenty years it has not been improved to the point of usefulness. (5) Possibly more radical suggestions would be worth considering. Suppose the test-by-test reviews in the vocational interest field, for example, were replaced by a single summary of the field, signed by Super, Bordin, Carter and Bennett. This could provide perspective, balance, and guidance far beyond that allowed by the present organization, however perfect MMY is by librarianship standards.

Few individuals have made so singular a contribution to the profession as Buros. Devotion, self-sacrifice, clarity of purpose, and editorial talent are evidenced in every page of Buros' series. It is gratifying that the leading measurement organizations have recently adopted a testimonial resolution recognizing this. Criti-

cism can say only that the world is perhaps not yet ready to support a project of this elegance, and that Buros can compromise with his editorial ideals without seriously compromising his professional contribution.

LEE J. CRONBACH

University of Illinois

PUBLICATIONS RECEIVED

- LOUIS O. KATTSOFF. *The Design of Human Behavior*. St. Louis (122 N. Seventh St.): Educational Publishers, Inc., 1953, pp. 402 \$5.00.
- ALFRED C. KINSEY, WARDELL B. POMEROY, CLYDE E. MARTIN, AND PAUL H. GEBHARD. *Sexual Behavior in the Human Female*. Philadelphia: W. B. Saunders Co., 1953, pp. 842. \$8.00.
- BRUNO KLOPPER, MARY D. AINSWORTH, WALTER G. KLOPPER, AND ROBERT R. HOLT. *Developments in the Rorschach Technique. Volume I: Technique and Theory*. Yonkers, N. Y.: World Book Co., 1954, pp. 726.
- HARVEY C. LEHMAN. *Age and Achievement*. Princeton, N. J.: Princeton University Press, 1953, pp. 359. \$7.50.
- CLARENCE LEUBA. *The Natural Man*. Garden City, N. Y.: Doubleday & Co., Inc., 1954, pp. 70 (paper) \$0.95.
- FRANCES LLOYD. *Educating the Sub-normal Child*. New York: Philosophical Library, 1953, pp. 148. \$3.75.
- WESLEY P. LLOYD. *Student Counseling in Japan: A Two-Nation Project in Higher Education*. Minneapolis: The University of Minnesota Press, 1953, pp. 204. \$4.00.
- VIKTOR LOWENFELD. *Your Child and His Art*. New York: The Macmillan Co., 1954, pp. 186. \$6.50.
- ROBERT WILLIAM LUNDIN. *An Objective Psychology of Music*. New York: The Ronald Press Co., 1953, pp. 303. \$4.50.
- ALBERT LYND. *Quackery in the Public Schools*. Boston: Little, Brown & Co., 1953, pp. 282. \$3.50.
- GORDON N. MACKENZIE, STEPHEN M. COREY, AND ASSOCIATES. *Instructional Leadership*. New York: Bureau of Publications, Teachers College, Columbia University, 1954, pp. 209. \$3.25.
- KARL MANNHEIM. *Essays on Sociology and Social Psychology*. New York: Oxford University Press, 1953, pp. 319. \$6.50.
- JAMES CLARK MOLONEY. *Understanding the Japanese Mind*. New York: Philosophical Library, 1954, pp. 252. \$3.50.
- MARGARET NAUMBURG. *Psychoneurotic Art: Its Function in Psychotherapy*. New York: Grune & Stratton, Inc., 1953, pp. 148. \$6.75.
- ADOLF NICHTENHAUSER, MARIE L. COLEMAN, AND DAVID S. RHUE. *Films in Psychiatry, Psychology, and Mental Health*. New York: (10 Downing St.): Health Education Council, 1953, pp. 269. \$6.00.

- CHARLES E. OSGOOD. *Method and Theory in Experimental Psychology*. New York: Oxford University Press, 1953, pp. 800. \$10.00.
- A. K. C. OTTAWAY. *Education and Society: An Introduction to the Sociology of Education*. New York: Grove Press (59 W. 9th St.), 1953, pp. 182.
- DONALD G. PATERSON, C. d'A GERKEN, AND MILTON E. HAHN. *Revised Minnesota Occupational Rating Scales*. Minneapolis, Minn: University of Minnesota Press, 1953, pp. 85. (paper) \$2.00.
- LEIGH PECK. *Child Psychology: A Dynamic Approach*. Boston: D. C. Heath & Co., 1953, pp. 536. \$5.25.
- JEAN PIAGET. *The Child's Conception of Number*. New York (55 E. 11th St.): The Humanities Press, Inc., 1953, pp. 248. \$5.00.
- EDWARD PODOLSKY. *The Jealous Child*. New York: Philosophical Library, 1954, pp. 147. \$3.75.
- EDWARD PODOLSKY, Editor. *Music Therapy*. New York: Philosophical Library, 1954, pp. 335. \$6.00.
- Progress of Literacy in Various Countries: A Preliminary Statistical Study of Available Census Data since 1900*. Monographs on Fundamental Education—VI. Paris: UNESCO, 1953, pp. 254. (paper)
- H. H. REMMERS. *Introduction to Opinion and Attitude Measurement*. New York: Harper & Brothers, 1954, pp. 437. \$5.00.
- JOHN W. M. ROTHNEY. *The High School Student: A Book of Cases*. New York: The Dryden Press, 1953, pp. 271. (paper) \$1.90.
- HELEN D. SARGENT. *The Insight Test: A Verbal Projective Test for Personality Study*. New York: Grune & Stratton, Inc., 1953, pp. 276. \$6.75.
- JEROME M. SEIDMAN, Editor, *The Adolescent: A Book of Readings*. New York: The Dryden Press, 1953, pp. 798. \$4.50.
- IRVING SETTEL, NORMAN GLENN AND ASSOCIATES. *Television Advertising and Production Handbook*. New York: Thomas Y. Crowell Co., 1953, pp. 480. \$6.00.
- GEORGE SIMPSON. *Man in Society: Preface to Sociology and the Social Sciences*. Garden City, N. Y.: Doubleday & Co., Inc., 1954, pp. 90. (paper) \$0.95.
- HERBERT SORENSON. *Psychology in Education*. Third Edition. New York: McGraw-Hill Book Co., 1954, pp. 577. \$5.50.
- HAROLD SPEARS. *Improving the Supervision of Instruction*. New York: Prentice-Hall, Inc., 1953, pp. 478. \$4.75.

- WILLIAM STEPHENSON. *The Study of Behavior: Q-Technique and Its Methodology*. Chicago: University of Chicago Press, 1953, pp. 376. \$7.50.
- W. S. TAYLOR. *Dynamic and Abnormal Psychology*. New York: American Book Co., 1954, pp. 658. \$5.50.
- R. MURRAY THOMAS. *Judging Student Progress*. New York: Longman, Green & Co., 1954, pp. 421. \$4.50.
- GODFREY. H. THOMSON, Chairman. *Social Implications of the 1947 Scottish Mental Survey*. London: University of London Press, 1953, pp. 356. 10s, 6d.
- JOHN C. TOWNSEND. *Introduction to Experimental Method for Psychology and the Social Sciences*. New York: McGraw-Hill Book Co., 1953, pp. 220. \$4.00.
- E. J. TRUEBLOOD. *The Dawn of the Post-Modern Era*. New York: Philosophical Library, 1954, pp. 198. \$3.75.
- LEONA E. TYLER. *The Work of the Counselor*. New York: Appleton-Century-Crofts, Inc., 1953, pp. 323. \$3.00.
- The Use of Vernacular Languages in Education*. Monographs on Fundamental Education VIII. Paris: UNESCO, 1953, pp. 154. (paper) \$1.00.
- CLYDE B. VEDDER. *The Juvenile Offender: Perspective and Readings*. Garden City, N. Y.: Doubleday and Co., Inc., 1954, pp. 510. \$6.00.
- PHILIP E. VERNON. *Personality Tests and Assessments*. New York (122 E. 55th St.): The British Book Centre Inc., 1953, pp. 220. \$4.00.
- HELEN M. WALKER AND JOSEPH LEV. *Statistical Inference*. New York: Henry Holt & Co., 1953, pp. 510. \$6.25.
- PAUL WITTY. *How To Become a Better Reader*. Chicago: Science Research Associates, Inc., 1953, pp. 216.
- NELLY WOLFFHEIM. (Translated by Charles L. Hannam.) *Psychology in the Nursery School*. New York: Philosophical Library, 1953, pp. 144, \$3.75.
- ARTHUR WORMHOUDT. *The Muse at Length: A Psychoanalytical Study of the Odyssey*. Boston: Christopher Publishing House, 1953, pp. 159. \$2.75.
- ORIN ROSS YOST. *What You Should Know about Mental Illness*. New York: Exposition Press, 1953, pp. 165. \$3.50.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 45

October, 1954

Number 6

READING ABILITIES OF BRIGHT AND DULL CHILDREN OF COMPARABLE MENTAL AGES¹

EMERY P. BLIESMER

University of Texas

In a recent study by Kolstoe (2), the usefulness of mental age as a unit of measurement was investigated by comparing the mental abilities of bright and dull children of similar mental ages. In general, his results revealed only minor differences in the traits measured. The present investigation is a companion study which involved essentially the same samples of bright and dull children as those employed by Kolstoe. Its purpose was to determine the extent to which children of equal mental age but markedly different in chronological age and IQ tend to be alike with respect to achievement in reading. This was done by comparing bright and dull children with approximately equal estimated true mental ages with respect to each of several abilities involved in reading comprehension. The criterion for 'dull' was an estimated true Stanford-Binet IQ of 84 or below; the one for 'bright' was an estimated true Stanford-Binet IQ of 116 or above. Thus the minimum separation between the two groups was approximately two standard deviation units along the IQ continuum. Estimates of the true IQ's and MA's were obtained by applying the standard formula for correction for regression to the obtained IQ's and MA's, with the reliability coefficients used in the formula being those reported by Terman and Merrill for various IQ levels.²

Results of the extensive investigations of Lewis (3) and of

¹ This article is based on the writer's Ph.D. dissertation. The work was done at the State University of Iowa under the co-direction of Professors A. N. Hieronymus and J. B. Stroud.

² L. M. Terman and M. A. Merrill, *Measuring Intelligence*, Boston: Houghton Mifflin Co., 1937, p. 46.

McGehee (4), earlier studies by Almack and Almack (1) and by Van Wagenen (8), and a later study by Thomas (6), suggest that, in general, bright children achieve below, and dull children achieve above, levels consistent with their indicated mental ages. One might expect, therefore, that a direct comparison of the achievement of bright and dull children with approximately the same mental ages would show results favoring the latter. The fact that the older, low IQ children have been in school longer and have been exposed to instruction in certain specific reading skills not taught in the lower grades would also lend support to such an hypothesis. On the other hand, most of the investigations of the relationship between brightness and over- and under-achievement have failed to take into consideration regression effects resulting from lack of reliability in the mental age measures, which would account for some of the observed differences. Ceiling and floor effects, and procedures used in extrapolating age and grade scales on standard reading tests may also be responsible for finding very little over-achievement among high IQ groups and under-achievement among groups of low IQ pupils. Thus, assuming equal variability of reading age scores and mental age scores, it is virtually impossible for a pupil at or near the 99th percentile on an intelligence test to 'over-achieve' or a pupil near the first percentile to 'under-achieve'.

In a study by Ramaseshan (5), ninth-grade pupils of similar mental ages (*Chicago Tests of Primary Mental Abilities*) were grouped according to chronological age. Children with similar mental ages but different chronological ages (and, consequently, differing IQ's) were compared on the subtests of the *Iowa Tests of Educational Development*. Her bright group was found to excel on all of the subtests of the ITED, but differences were statistically significant for only three of these subtests. Of some pertinence to the purposes of the writer's study is the fact that no significant differences were found on the three subtests measuring the ability to read and interpret material in specific content areas.

Unsicker (7) equated groups according to mental age by matching groups of bright third- and fourth-grade pupils with groups of dull seventh- and eighth-grade pupils, first on the basis of *Kuhlman-Anderson Intelligence Test* scores, then on the basis of *California Tests of Mental Maturity* results. When the groups matched on the basis of Kuhlman-Anderson MA's were compared

with respect to reading comprehension scores earned on the *Iowa Tests of Basic Skills*, Unsicker found significant differences favoring the bright group. Differences between the groups matched on the basis of California MA's were not significant.

In addition to the fact that different criterion tests were employed, the present study differs from most of those in which the problem has been investigated previously in that estimated true scores were used in the matching procedure, individually administered intelligence tests rather than group tests were employed, and the criterion tests were of an appropriate level of difficulty so that ceiling and floor effects did not operate to bias the means.

PROCEDURE

Method of Sampling.—In order to measure a wide range of comprehension abilities, it was decided to employ test materials which are appropriate at the fourth- to fifth-grade level of reading ability. Considering the IQ criteria for 'bright' and 'dull' (lower and upper IQ limits of 116 and 84, respectively), and studying the overlap of MA's at various CA's for these bright and dull criteria, it was decided to use a mental age range of from ten years, seven months, through twelve years, six months, and to restrict the bright group to children with CA's of ten years or less and the dull group to children with CA's of fourteen years or above.

Children in the dull group were selected from regular eighth- and ninth-grade classes and some special education classes in two junior high schools, and children in the bright group from regular third- and fourth-grade classes in three elementary schools, in the public school system of a large Iowa city. For identification of pupils likely to meet sample specifications, the cumulative record folder of each child in the two junior high schools, and in third and fourth grades in the three elementary schools, was studied. A list was made of all the junior high pupils' who were fourteen years of age or older and for whom IQ's of 90 or below had been obtained with group intelligence tests which had been administered by the schools in previous years. In the elementary schools, a list was made of all third- and fourth-grade pupils who were ten years of age or younger and for whom IQ's of 110 or above had been obtained with group intelligence tests in previous years.

From the list of pupils indicated as likely to meet specifications for the dull group, pupils were selected randomly and the *Revised*

TABLE I.—CHARACTERISTICS OF THE SAMPLES

	Groups	
	Bright	Dull
Range of CA's	8-7 through 9-10	14-2 through 16-3
Mean CA	9-2.5	15-5.4
Range of MA's (est. true)	10-8 through 12-6	10-8 through 12-6
Mean MA	11-3.2	11-3.0
Range of IQ's (est. true)	116 through 138	72 through 84
Mean IQ	126.5	79.5

Stanford-Binet Scale, Form L, was administered until there were obtained twenty-nine pupils who met the following specifications:

- 1) Estimated true IQ's of 84 or below.
- 2) Estimated true MA's of from ten years, seven months, through twelve years, six months.

From the list of third- and fourth-grade pupils indicated as likely to meet sample specifications for the bright group, pupils were selected randomly and the *Revised Stanford-Binet Scale*, Form L, was administered until there was obtained a sample of pupils meeting the following specifications:

- 1) Estimated true IQ's of 116 or above.
- 2) Estimated true MA's of from ten years, seven months, through twelve years, six months.
- 3) As many bright children in each of four six-months mental age intervals or levels (which constituted the two-year range indicated in the preceding specification) as there were dull children in that interval.

In the process of obtaining enough bright cases in each level to match the number of dull cases in that level, extra cases were obtained for some of the levels. For purposes of statistical analysis, extra cases in each level were later discarded randomly. The final sample contained twenty-eight children in each group, dull and bright, one case originally selected for the dull group having been lost because of incomplete data. In the dull group, there were fifteen boys and thirteen girls; thirteen in Grade 8, six in Grade 9, and nine in special classes. The bright group was composed of sixteen boys and twelve girls; fifteen in Grade 3 and

thirteen in Grade 4. A summary of information relative to the dull and bright groups in the sample is presented in Table I.

Selection of Comprehension Abilities Investigated.—A survey of the professional literature, teachers' manuals accompanying series of readers, standardized tests, and reported results of factorial studies was made for suggestions of specific abilities involved in reading comprehension. Consideration was also given to availability of measuring instruments for these specific abilities, the possibilities of adapting available instruments to the purposes of this study, and the purported importance of given abilities. This resulted in the selection of the following abilities for inclusion in this investigation:

- 1) Word Recognition (the ability to recognize given words 'on sight').
- 2) Word Meaning (the ability to understand or recognize the particular meaning of a word as it is used in context).
- 3) Memory for Factual Details (the ability to recall specific facts which have been definitely stated in a selection).
- 4) Location or Recognition of Factual Details (the ability to locate or recognize specific factual details which are explicitly stated in a given selection).
- 5) Perception of Relationships among Definitely Stated Ideas (the ability to recognize or to formulate an idea which is not explicitly stated in a selection but which is contained in the selection when two or more definitely stated ideas are considered together).
- 6) Recognition of Main Ideas (the ability to recognize the central thought or main idea of an entire selection, a paragraph, or a specific part of a paragraph).
- 7) Drawing Inferences and Conclusions (the ability to recognize, or to formulate, an idea which is not stated in a selection but which is dependent upon the combination of an idea (or ideas) which is (are) definitely stated in a selection and one which is outside the selection and within the informational or experiential background of the individual).

In addition to these abilities, measures of reading rate and listening comprehension were also obtained. Reading rate scores were secured not only because comparisons between the groups were of some interest in themselves, but also because marked

differences in rate could be responsible for differences in comprehension abilities, even though the tests were untimed. The measures of listening comprehension were obtained in order to determine whether differences in general comprehension ability exist when unencumbered by possible difficulties with mechanical skills in reading.

Criterion Tests Used.—Eighty words from the Flashed Word Recognition and Word Analysis Test of the *Durrell Analysis of Reading Difficulty*, and twenty more difficult words from various forms (Q through T) of Part II (Vocabulary) of the Reading Comprehension Test, Advanced Battery, of the *Iowa Tests of Basic Skills*, were included in the Word Recognition Test. These one hundred words were arranged in lists of twenty each and were presented tachistoscopically for a duration of approximately one second. Only results obtained with the last four lists (eighty words) were included in the analysis of results.

The Word Meaning Test was a multiple response test consisting of a representative sample of fifty items chosen from various forms (Q through T) of Part II (Vocabulary) of the Reading Comprehension Test, Elementary and Advanced Batteries, of the *Iowa Tests of Basic Skills*.

A test of Comprehension Abilities was made up of nine reading selections and one hundred and thirty items which were adapted from the Reading Comprehension Test of the Elementary and Advanced Batteries of the *Iowa Tests of Basic Skills*, Forms L through T. Five subtest scores, each based on from twenty-five to twenty-seven items, were obtained as measures of these five specific abilities: Memory for Factual Details, Location or Recognition of Factual Details, Perception of Relationships among Definitely Stated Ideas, Recognition of Main Ideas, and Drawing Inferences and Conclusions. The reading selections were chosen on the basis of relative difficulty, apparent interest value, and the extent to which items accompanying selections represented the specific abilities named above. To obtain subtests of approximately equal length and difficulty, a number of original items were eliminated and additional ones were constructed when necessary.

A Listening Comprehension Test was constructed in a manner similar to that for the Test of Comprehension Abilities, except that items for Location or Recognition of Factual Details were not included. The test consisted of four reading selections and forty-

two items. The test booklets contained only the questions related to the selections. The selections and the test items were read aloud to the subjects by the examiner, and subjects marked their chosen responses to each question after it and its answer choices had been read.

The Reading Rate Test consisted of a selection of approximately fifteen hundred words of fourth- to fifth-grade level of difficulty. Subjects were instructed to read the selection once "at the same speed as you usually read." They were told before they began that there would be questions over the material read. A short comprehension test, consisting mainly of items measuring memory for details, followed the reading of the selection. The rate score was a complement of the number of complete ten-second intervals which had elapsed during the reading of the selection.

All of the criterion tests were administered by the writer during an eight-day testing period. All were administered as group tests, without time limits, with the exception of the Word Recognition Test, which was administered to each subject individually. All test items, except those for the Word Recognition Test, were of the four-choice multiple-response type. Except for the Reading Rate Test rate scores, all scores used in the analyses of results represented the number of items answered correctly.

Analysis of Results.—An analysis of variance design, 'group-by-levels', was employed in the analysis of results.³ The bright and dull children represent the 'groups,' and the intervals of six months in the two-year mental age range used in the study constitute the 'levels.' A schematic presentation of the groups-by-levels design as it applies to this study is shown below.

MA Level	MA (Es'd True)	Criterion Scores	
		Bright Group	Dull Group
I	12-1 to 12-6	n = 7	n = 7
II	11-7 to 12-0	n = 9	n = 9
III	11-1 to 11-6	n = 7	n = 7
IV	10-7 to 11-0	n = 5	n = 5

In the case of each ability investigated, the null hypothesis was

³ E. F. Lindquist, *Design and Analysis of Experiments in Psychology and Education*. Boston: Houghton Mifflin Co., 1952.

TABLE II.—SUMMARY OF RESULTS: MEAN SCORES FOR EACH GROUP, DIFFERENCES BETWEEN MEANS, AND F-VALUES, FOR EACH ABILITY

Ability	Bright		Dull		Differences ($M_B - M_D$)	F*
	Mean	SD	Mean	SD		
Word Recognition	58.6	11.8	56.2	13.0	+2.4	0.499
Word Meaning	30.4	6.1	30.8	6.1	-0.4	0.076
Memory for Factual Details	17.4	4.0	15.1	4.8	+2.3	3.678
Location or Recognition of Factual Details	15.8	3.6	12.7	3.3	+3.1	10.126**
Perception of Relationships among Definitely Stated Ideas	15.4	3.6	14.0	3.6	+1.4	2.058
Recognition of Main Ideas	16.3	4.1	13.9	3.9	+2.4	5.319**
Drawing Inferences and Conclusions	14.9	3.1	12.9	3.7	+2.0	4.628**
Total Comprehension Abilities	79.8	15.1	68.4	16.7	+11.4	7.922**
Listening Comprehension	30.6	3.6	23.6	4.7	+7.0	47.574**
Reading Rate (rate score)	61.6	17.9	55.4	15.9	+6.2	1.602
Reading Rate (comprehension score)	10.7	2.4	9.8	3.4	+0.9	1.444

* For each F-value, $df = 1, 48$; $F_{05} = 4.04$.

** Significant at 5 per cent level.

tested, *i.e.*, that the means of the populations of which the dull and bright groups were representative samples were the same. To test this hypothesis, the ratio of the mean square for groups to the mean square for within cells was employed. This ratio yields a value which is distributed as F, provided that the hypothesis is true and that certain conditions are met. A five per cent coefficient of risk, selected in advance of the analyses, was employed in rejecting the null hypothesis.

RESULTS

A summary of the obtained results is presented in Table II. The table includes the mean and standard deviations for each group, the differences between the means, and the F-values obtained in the tests of the significance of the differences between

the means of the bright and dull groups. Positive differences favor the bright group. .

Obtained differences between the mean scores of the two groups favored the bright group with respect to all of the abilities except Word Meaning. In the case of this one exception, the difference was not only nonsignificant, but also less than one raw score unit. The analyses of variance yielded significant differences with respect to the following five abilities: Location or Recognition of Factual Details, Recognition of Main Ideas, Drawing Inferences and Conclusions, Total Comprehension Abilities, and Listening Comprehension. While not significant at the required level, differences between the mean scores for Memory for Factual Details and for Perception of Relationships among Definitely Stated Ideas were substantial, and further investigation with respect to these two abilities seems warranted. Differences with respect to Word Recognition, Word Meaning, and Reading Rate were not significant.

The Test of Listening Comprehension was included in anticipation of the possible event that the group which was found significantly poorer on most of the abilities would be found to be significantly better with respect to listening comprehension. Then such results might have been interpreted in terms of possible difficulty with mechanical skills in reading rather than in terms of differences in intellectual abilities or specific comprehension abilities. However, the bright group, which excelled on nearly all of the abilities, also excelled significantly on the Listening Comprehension Test. Similarly, if a significantly higher rate had been found for the group also found to be significantly poorer with respect to most of the abilities, the poorer showing might have been attributed, in part, to tendencies to read carelessly and too hastily. However, no significant rate difference was found between the two groups and analysis of the reading rate comprehension check indicated that the two groups had read the reading rate selection with comparable degrees of understanding. These findings with respect to reading rate and listening comprehension tend to further indicate that superiority in reading comprehension involves superiority in intellectual functions rather than in the more mechanical skills.

Tests for interaction between groups and levels were also made.

No significant interaction effects were found for any of the abilities tested, thus satisfying one of the necessary conditions or underlying assumptions involved in the particular design of the study.

Inspection of the frequency distributions for the various abilities involved in this study revealed that, in general, obtained scores did not closely approach the maximum possible at the upper end of the distribution or the 'chance' scores at the lower end. Thus, neither ceiling nor floor effects operated to bias the results.

CONCLUSIONS

In the strictest sense, the sample studied may be regarded as a representative sample only of hypothetical populations that show the same relative distribution of MA's, IQ's, and CA's as the groups in the sample itself; and generalizations based upon obtained results should be restricted to these hypothetical populations. However, since no significant interaction effects between groups and levels were found, restrictions upon extending generalizations to real populations may be lifted to a considerable extent. Therefore, generalizing to a population of dull and bright children with widely differing IQ's but approximately equal MA's within the MA ranges found in this study and with reference to the various comprehension abilities as defined operationally, the following conclusions seem warranted:

- 1) Bright children are significantly superior to dull children of comparable mental ages with respect to total reading comprehension and the following specific abilities: locating or recognizing factual details, recognizing main ideas, and drawing inferences and conclusions.

- 2) It seems probable that bright children are also superior to dull children of comparable mental ages with respect to memory for factual details and perception of relationships among definitely stated ideas.

- 3) Bright children are superior to dull children of comparable mental ages with respect to listening comprehension.

- 4) Reading rates of bright and dull children of comparable mental ages appear to be approximately the same when comparable degrees of understanding of material read are attained, with a wide range in rate being found in both groups.

- 5) Bright and dull children tend to be alike with respect to ability in word recognition and word meaning. Bright children

are significantly superior to dull children of comparable mental ages with respect to the relatively more complex, and intellectual, comprehension abilities.

6) It would seem that levels of expectation with respect to the more complex comprehension abilities should not be as high for dull children as for bright children of comparable mental ages.

REFERENCES

- 1) John C. Almack, and James L. Almack. "Gifted pupils in the high school." *School and Society*, 14: 227-8, September 24, 1921.
- 2) Oliver P. Kolstoe. "A comparison of mental abilities of bright and dull children of comparable mental ages." *Journal of Educational Psychology*, 45: 161-68, March, 1954.
- 3) W. Drayton Lewis. *A Study of Superior Children in the Elementary School*. George Peabody College Contributions to Education, No. 266, George Peabody College for Teachers, Nashville, Tenn., 1940.
- 4) William A. McGehee. *A Study of Retarded Children in the Elementary School*. George Peabody College Contributions to Education, No. 246, George Peabody College for Teachers, Nashville, Tenn., 1939.
- 5) Rukmini S. Ramaseshan. "A note on the validity of the mental age concept." *Journal of Educational Psychology*, 41: 56-58, January, 1950.
- 6) G. I. Thomas. "A study of reading achievement in terms of mental ability." *Elementary School Journal*, 47: 28-33, September, 1946.
- 7) Willard D. Unsicker. *A Psychological Study of Bright and Dull Children with Comparable Mental Ages*, unpublished Doctor's dissertation, State University of Iowa, Iowa City, Iowa, 1950.
- 8) M. J. Van Wagenen. "A comparison of the mental ability and school achievement of the bright and dull pupils in the sixth grade of a large school system." *Journal of Educational Psychology*, 16: 186-92 March, 1925.

GROUP PROCEDURES MODIFYING ATTITUDES OF PREJUDICE IN THE COLLEGE CLASSROOM¹

GERALD S. WIEDER

Brooklyn College

Community Counseling Center

Since recent research has indicated the intimate relationship between prejudice and certain patterns of emotional development and expression, it would appear logical that one approach to modifying prejudice might be through techniques which focus primarily on many of the same emotional factors. Thus, within the setting of the college classroom an experiment was designed which would compare the relative effectiveness of two methods of instruction in modifying attitudes associated with racial, religious and ethnic prejudice: one method utilized such group therapy procedures as nondirective methods of instruction and sociodrama, while the other employed the traditional lecture-discussion method. A secondary purpose of this research was to investigate the relationship between personality adjustment and the modification of the social attitudes with which this study is concerned. Although not directly within the scope of this investigation, the subsidiary problem of course content acquisition for the two classes taught by the investigator was examined.

Recent research on the problem of minority-group prejudice has indicated the close relationship between certain socially-determined personality variables and prejudice. It has also been demonstrated that client-centered and sociodramatic therapeutic procedures can be adapted to classroom situations. (2, 3, 4) Consequently, it would appear logical that one approach in modifying prejudice might be through utilizing group techniques in a classroom situation which focus primarily on many of the same personality variables.

¹ This article is a summary of a Doctoral dissertation completed at New York University. It was sponsored by Professor Charles E. Skinner, to whom the investigator wishes to express his deep appreciation for the interest and assistance provided.

The experimental design was organized in a manner which would permit integration of the research techniques into the fabric of the courses. Three instruments were employed; namely, the California Public Opinion Study E-F Scales, the Index of Adjustment and Values, and the Minnesota Multiphasic Inventory, both before and after the experimental period, to measure any changes. An objective-type course-content examination prepared by the investigator was administered to the two experimental classes at the midpoint and ending of the semester. The experimental classes were conducted by the investigator, who randomly selected one group to be taught by the utilization of group therapy procedures and the other to be conducted by the traditional lecture-discussion method. A third class, taught by another instructor using traditional methods of instruction, served as one control. Finally, another group which was composed of students registered in various courses within the same curriculum served as a time-factor control. All groups were compared on the significant variables and appeared to come from the same parent population.

THE PROCEDURE

The actual steps taken in the organization of this research are as follows:

A) In September, approximately one week prior to the beginning of the school semester, the California E-F Attitude Scales, the Minnesota Multiphasic Personality Inventory, and the Otis Self-Administering Tests of Mental Ability were administered to a group of students (Control Group I) entering the Vocational Diploma Program at Brooklyn College.

B) During the second class session of the two sections of Practical Psychology D1 (the group therapy class and the traditional lecture-discussion class), given by the investigator, and of a third section of the same course given by another instructor (Control Group II), the E-F Attitude Scales and the Minnesota Multiphasic Personality Inventory were administered. The Otis Self-Administering Tests of Mental Ability was part of the entrance testing battery of all Vocational Diploma Students desiring admission to the College.

C) The investigator conducted one section by the traditional lecture-discussion method and the other using group-therapy

procedures. The cooperating instructor conducted his section by the traditional lecture-discussion method. The fourth group was distributed among many classes, many of the students not registered for Practical Psychology D1.

D) During the seventh class meeting the Bills' Index of Adjustment and Values was administered to the two classes conducted by the investigator.

E) During the last session of the semester the four groups were retested with the original test instruments, including the Index of Adjustment and Values, with the exception of the Otis Self-Administering Test of Mental Ability.

Before- and after-testing of attitudes and personality was conducted in order to measure any changes which might have taken place as a result of the teaching methods used. The purpose of the control groups was to estimate the effect of factors other than those of the teaching methods used; i.e., bias of instructor and the many uncontrollable variables attendant on just being registered in college for a similar period of time. Thus, there were two kinds of comparisons possible to evaluate the effects of the educational methods used. First, an intra-group comparison consisting of noting differences in the same group at the beginning and end of the semester; and, second, the inter-group comparison of noting the differences among the four groups studied.

Since any conclusions based upon this investigation are dependent in good part on the experimental variables introduced, it would seem necessary to offer some description of the differing methods used in treating the subject matter of the course.

TRADITIONAL LECTURE DISCUSSION METHOD

The procedure and plan of instruction followed with the lecture-discussion section is familiar to most college graduate and undergraduate classes. The instructor had rather carefully-planned lecture material which he presented to the class in a more or less systematically directed fashion towards the end of the covering of a given amount of pertinent factual material. Student participation was usually limited to isolated questions and even more isolated comments regarding the factual material presented. A great deal of note-taking was commonplace, with the usual cram sessions for mid- and end-term examinations. The following are representative class materials presented:

Third Meeting.—Heredity-environment including genetic approach, cultural-anthropological approach, history of controversy, studies, present thinking.

Twelfth Meeting.—Husband-wife relationships including authority and responsibility, the career woman, economic factors, in-laws, physiological factors and divorce.

GROUP THERAPY METHOD

The type of play drama used with this class was a modification of the psychodramatic and sociodramatic sequences developed and used by Moreno (5) and Jennings (4). During the first class meeting the students were asked to list five everyday situations either at home or school that caused them embarrassment, created tension, or evoked displeasure, anger or unhappiness. The students were then asked to indicate whether they would care to appear in a little play built around this situation which would be presented to the group.

Many of the dramas selected for class use were drawn from those presented by the students. The criterion for selection, of necessity, was their applicability to the material of the course. Where there were none which would adequately touch on course content, the investigator created one appropriate to the topic of discussion.

Each play-drama was structured by the instructor, but the students were encouraged to select their own rôles and play them out spontaneously, making up lines as they went along. The procedure usually followed was to permit the players to terminate their own scenes, with a half-hour being arbitrarily selected as a maximum time limit. Following the drama, the 'audience' was then invited to comment on any phase of the drama that to them invited comment. Concentrated effort was made to keep the two courses fairly similar in terms of subject matter treated and in the amount of time devoted to each area covered. Hence the subject of prejudice itself was confined to one class session in each class.

Examples of specific play-dramas used in connection with the development of course content paralleling the examples selected from the traditionally conducted class are as follows:

Heredity-Environment.—A young couple who have been seeing each other for the past year decide that they are deeply in love and wish to marry. The fly in the ointment, however, is that the boy's father is serving time in the federal penitentiary for a crimi-

nal offense. The young man's conduct and behavior have always been exemplary, but the young lady is aware of the strong objections of her parents to this contemplated marriage. They nevertheless decide to have the matter out with the girl's parents that night.

Husband-Wife Relationships.—a) The younger brother never seems to be able to get dressed early enough to get to school on time. Father, in a fit of anger at this recurrent situation begins to argue with mother. b) Mother, unable to cope with the children asks father to do something about it. Father answers by declaring it's not his responsibility.

RESULTS

Table I presents the mean attitude scale scores, variability and *t* ratios of the therapy groups at the beginning and end of the semester. There were three attitude scale variables on which the significance of changes were tested. These were the E, F, and E-F scales. The E-F score was the combined prejudice score as measured by the entire attitude scale.

Inspection of Table I reveals that differences obtained from pre- to post-testing were significant for the group therapy class. More specifically, changes occurring on the E scale might be

TABLE I.—MEANS, STANDARD DEVIATIONS AND COMPARISONS OF PRE- AND POST-TESTS ON THE E, F AND E-F SCALES FOR THE THERAPY GROUP

Scale	N	Pre-test		Post-test		$M_1-M_2^1$	F	"t"	Level of Significance
		M	σ	M	σ				
E	27	65.4	13.04	60.9	15.1	4.22	1.33*	2.06	.05
F	27	122.1	23.36	105.4	22.7	15.92	1.06*	5.02	.01
E-F	27	187.1	31.57	167.1	30.0	20.46	1.10*	6.46	.01

NOTE: Using the small sample techniques and entering the "t" table with 26, 27, and 28 (as the case may be) degrees of freedom the ratio is significant at the .05 level if it is over 2.00 and at the .01 level if it is over 2.65 (H. Garrett, *Statistics in Psychology and Education*, New York: Longman, Green and Co., 1947, p. 190).

¹ The difference between each pre- and post-test score was calculated for each person on each item by subtracting the latter from the former. Means and standard deviations of these differences were then compared.

* F is not significant at the .05 level of confidence.

expected, due to chance, in only five out of one hundred cases: changes in the F scale in only one out of a hundred cases, and for the combined scale significance is also well established in that such changes as occurred would probably occur by chance in only one out of one hundred cases. Inspection of F results reveals no significant difference in variability.

Such changes as did occur on the E, F and combined E-F scales for the lecture-discussion, Control Groups I and II were not statistically significant. With the exception of Control Group II, F and combined E-F scales, the direction of change was towards a reduction of prejudice. Changes in variability were likewise not significant.

Table II presents means, variability and t ratios on the Index for the experimental group at the beginning and end of the semester. In the therapy group, Scale I, which is an expression of the individual self-concept, increased significantly in that the changes that occurred could be attributed to chance factors but five times out of one hundred. Scale II, which is an expression of the accepting or rejecting nature of the feelings the individual holds towards self, increased even more significantly in that the differences were significant at the .01 level of confidence. Scale III, representing the expression of the ideal self, was raised but not to any degree of statistical significance. Scale III-I which Bills has termed the discrepancy index and which relates to the discrepancy between the individual's ideal self and his self-concept has been described

TABLE II.—MEANS, STANDARD DEVIATIONS AND COMPARISONS OF PRE- AND POST-TESTS ON THE INDEX OF SELF ADJUSTMENT FOR THE THERAPY GROUP

Scale	N	Pre-test		Post-test		$M_1-M_2^1$	F	"t"	Level of Significance
		M	σ	M	σ				
I	27	178.67	14.38	184.61	13.91	6.09	1.06*	2.11	.05
II	27	189.17	21.84	202.00	18.90	13.33	1.33*	3.76	.01
III	27	194.44	16.35	196.46	13.71	3.72	1.42*	1.48	N.S.
III-I	27	15.72	15.10	13.11	10.27	-1.18	2.16*	.51	N.S.

¹ The difference between each pre- and post-test score was calculated for each person on each item by subtracting the latter from the former. Means and standard deviations of these differences were then compared.

* F is not significant at the .05 level of confidence.

as an index of maladjustment. Positive movement in psychotherapy should be accompanied by reduction in the discrepancy index. Results for the group therapy class indicate this decrease in scale III-I but not to any degree of statistical significance.

The data obtained for the lecture-discussion group reveal that such changes as were effected did not attain or approach statistical significance. Although the means of Scale I did increase somewhat, means for Scales II and III decreased. The Discrepancy Index moved in the direction of better adjustment but did not approach significance.

As in the lecture-discussion group the data for Control Group I indicate that changes effected were not statistically significant. Movement on Scales I, II and III were all in the direction of decreased means representing a decline in feelings of self-worth, self-acceptance, and of ideal self. Movement in the Discrepancy Index was in the direction of greater adjustment, but did not approach statistical significance.

The data obtained for Control Group II reveal that with the exception of Scale III changes effected did not approach statistical significance. Movement on Scales I and II was in the direction of a decrease in self-worth and self-acceptance. Significant, however, at the .01 level of confidence is the movement of this group in the direction of a lowered ideal self. The discrepancy index was in the direction of greater adjustment but did not approach statistical significance.

Table III presents the intra-group changes from the beginning to the end of the semester in the experimental group along with the "t" ratios testing the significance of these changes. Only two of the twelve variables in the group therapy class reached significance at the .05 level of confidence, Pd (Psychopathic deviate) and M (Masculinity-Femininity) at the .01 and .05 levels, respectively, although the Pa (Paranoia) and the K (Correction factor) did approach significance. Table III also shows that the direction of movement on the L (Lie), Hs (Hypochondriasis), D (Depression), M (Masculinity-Femininity), Pt (Psychasthenia) and Sc (Schizophrenia) was toward decreased scores. Movement on the K (Correction factor), F (Validity scale), Hy (Hysteria), Pd (Psychopathic deviate), Pa (Paranoia) and Ma (Hypomania) was toward increased scores.

In the lecture-discussion class four of the twelve variables

TABLE III.—MEANS, STANDARD DEVIATIONS AND COMPARISONS OF PRE- AND POST-TESTS ON THE M.M.P.I. FOR EXPERIMENTAL CONTROL GROUPS. (A) GROUP THERAPY

Scale	Pre-test		Post-test		Change between Pre- and Post-test ¹		"t"	Level of Significance
	M	σ	M	σ	M	σ		
L	54.95	6.58	54.5	5.89	-.25	5.89	.21	N.S.
K	59.02	7.29	61.39	9.13	2.20	6.77	1.65	.20
F	53.22	4.67	53.61	6.70	.35	5.99	.29	N.S.
Hs	54.43	8.78	52.87	6.23	-1.13	8.02	.71	N.S.
D	53.89	10.05	52.33	9.50	-1.21	5.30	1.16	N.S.
Hy	55.11	9.15	57.09	6.75	1.83	6.49	1.43	N.S.
Pd	54.78	9.22	60.44	10.45	5.44	9.93	2.79	.01
M	52.94	6.36	49.54	8.19	3.45	7.20	2.41	.05
Pa	49.31	6.51	53.00	8.83	2.79	8.09	1.75	.10
Pt	54.28	10.38	52.94	8.75	-1.62	10.41	.79	N.S.
Sc	55.61	6.24	55.46	6.96	-.32	6.89	.23	N.S.
Ma	53.44	8.25	56.33	10.01	2.98	11.24	.14	N.S.

NOTE: Using the small sample techniques and entering the "t" table with 26, 27, and 28 (as the case may be) degrees of freedom the ratio is significant at the .05 level if it is over 2.00 and at the .01 level it is over 2.65.

¹ The difference between pre- and post-test scores was calculated for each person on each item by subtracting the latter from the former. Means and standard deviations of these differences were then compared.

revealed significance at the .05 level of confidence. These were in K (Correction factor) .01 level, Hs (Hypochondriasis) .05 level, Hy (Hysteria) .02 level and Sc (Schizophrenia) .05 level scales. The direction of movement of these was towards increased scores, as it was also on L (Lie), K (Correction factor), F (Validity scale), Pd (Psychopathic deviate) and Pa (Paranoia), though the latter were not significant. Decrease in mean scores occurred in the D (Depression), M (Masculinity-Femininity) and Ma (Hypomania), but these changes were likewise not statistically significant.

Only one of the twelve variables in Control Group I reached significance at the .05 level of confidence, the F (Validity scale). Movement towards a decrease in mean score occurred in this and also in the Hs (Hypochondriasis), D (Depression), Hy (Hysteria), M (Masculinity-Femininity), Pa (Paranoia), Pt (Psychasthenia), and Sc (Schizophrenia) scales. Movement towards increased

scores occurred in the L (Lie), K (Correction factor), Pd (Psychopathic deviate), and Ma (Hypomania).

None of the variables in Control Group II attained significance at the .05 level of confidence. Movement towards increased scores took place in the L (Lie), F (Validity scale), Hs (Hypochondriasis), Hy (Hysteria), M (Masculinity-Femininity), Pa (Paranoia), Pt (Psychasthenia), and Sc (Schizophrenia) scales. Movement towards a decrease in score occurred in the K (Correction), D (Depression), Pd (Psychopathic deviate), and Ma (Hypomania) scales.

In interpreting these data it should be noted that both pre- and post-testing scale means for all groups did not reach abnormality. Thus, while it is possible to discuss significant scale shifts in relation to pre-test scores, it must be stressed that there were actually no significant changes in total adjustment for any of the groups on any of the scales of the MMPI. The significance of scale shifts towards increased means can therefore be considered of extremely tentative significance with relation to the clinical entity each scale is designed to measure. Decreases in means can generally be disregarded in view of the small range of deviation.

The Group Therapy Class had as many shifts in mean scale scores in the direction of maladjustment as it had in the direction of greater adjustment. The Pd (Psychopathic deviate) scale, which purports to measure a disregard of social mores, moved significantly in the direction of decreased adjustment, while the M (Masculinity-Femininity) scale moved significantly in the opposite direction. A possible interpretation of the Pd (Psychopathic deviate) scale shift is that the group responded to the therapeutic climate by an increased tendency to flout social mores or conventions. A plausible speculation, in view of the fact that the group mean did not approach borderline significance on the psychopathic deviate scale, is that the movement may be interpreted as a shift away from conventional behavior and a decrease in the repressive aspects of super-ego motivation. The plausibility of this speculation is further supported by the F (Validity scale) results where it was seen that the group significantly declined in expression of rigidly-held conventional attitudes. Supportive, too, are the data with regard to the heightened self-concept and self-acceptance indices of this group. It is logical to assume that movement in this direction may be permissive of greater felt freedom to depart from the expected behavior pattern. As indicated above,

the decrease in F (Validity scale) means may, for all intents and purposes, be disregarded.

Scale results for the Lecture-Discussion class present a rather perplexing picture. Movement in the direction of pathological behavior occurred on the K (Correction), Hs (Hypochondriasis), Hy (Hysteria) and Sc (Schizophrenia) scales. Strictly interpreted, it would seem that this class responded to the lecture-discussion approach of the investigator by tending to move in the direction of greater defensiveness, somatic complaints, and withdrawal behavior. A highly tentative speculation concerns the dual rôle played by the investigator. It is conjectured that the possible effects of this double identity may have contributed to the generation of more than the usual amount of anxiety caused by classroom practice and that this was manifested by characteristic college student defenses such as increased defensiveness, somatic pre-occupation, and greater intellectualization.

With the exception of a significant increase in the F (Validity scale) for Control Group I there was no significant movement for both controls. The high F (Validity scale) would suggest that the results for this group may be of questionable validity.

Table IV presents the class means, deviations, and ranges for both groups taught by the investigator, e.g., group therapy and lecture-discussion, on both mid- and end-term objective examinations. Examination of Table IV clearly indicates that while the Lecture-Discussion group did slightly better on both mid- and end-term examinations these differences are statistically insignifi-

TABLE IV.—A COMPARISON OF CONTENT EXAMINATION RESULTS FOR GROUP THERAPY AND LECTURE-DISCUSSION CLASSES

Class	Mid-Term				End-Term			
	M	σ	R	"t"	M	σ	R	"t"
Group Therapy (33)	81.51	5.51	70-90	.18	80.97	7.74	70-91	.75
Lecture-Discus- sion (25)	81.76	4.50	68-96		82.40	6.47	70-93	

NOTE: Using the small sample techniques and entering the "t" table with 56 degrees of freedom the ratio is significant at the .05 level if it is over 2.00 and at the .01 level if it is over 2.65.

cant. It is interesting to note that grade scores remained constant from the beginning to the end of the semester.

The general conclusions drawn from a statistical analysis of the data of this study are as follows:

1) The major purpose of the study was accomplished in the finding that a course in practical psychology conducted by using such group therapy procedures as student-centered classroom practice and sociodrama measurably modified attitudes associated with racial, religious, and ethnic prejudice.

2) A course in practical psychology when conducted along traditional lecture-discussion methods of instruction did not significantly modify these same attitudes.

3) The factor of attendance at Brooklyn College within the Vocational Diploma Program for the experimental period did not produce significant modification of these same social attitudes.

4) Although statistically significant changes took place on both E and F scales for the experimental group only, statistical analysis of the data revealed that the attitude measured by the F scale (covert attitudes of prejudice) was modified more significantly.

5) Inspection of the variable changes comprising both E and F scales revealed that the most significantly effected scale clusters were the conventionality and the authoritarian aggression items. Scale variables approaching significance were the authoritarian-submission and the superstition and stereotype items.

6) The data on the Index of Adjustment and Values were generally supportive of the attitude-scale findings. Significant increases in the Self-Concept (Scale I) and Acceptance of Self (Scale II) took place for the Group-Therapy class only. The discrepancy index (adjustment index III-I) while moving in the direction of greater adjustment did not approach significance.

7) The Minnesota Multiphasic Inventory did not reflect readily understandable results. Significant movement took place within the Group Therapy class on the Pd (increased) and M (decreased) variables. The Lecture-Discussion class shifted significantly upward on the K, Hs, Hy, and Sc scales. Movement within the control groups was limited to a significant increase in the F variable for Control Group I. It appears that the MMPI when used with a population of 'normal' individuals may not be the ideal instrument to measure movement towards greater or lesser adjustment within a 'normal' frame of reference.

8) A subsidiary finding for the two groups taught by the investigator was the practical similarity of course content test results on both mid- and end-term examinations.

IMPLICATIONS

The results of the present study have rather clearly demonstrated the relative superiority of therapeutic classroom techniques in modifying significantly inter-group prejudice. In doing so it has supported the findings of those investigators who believe prejudice to be but a symptom of a personality need syndrome, which is not readily amenable to modification by factual information, exhortation, propaganda or other simple linear approaches. It further supports the view of such clinical educators, if such they may be called, as Rogers, Cantor, Snygg, Combs, Shedlin and Kilpatrick, to mention but a few, who insist that education is only truly effective in the degree that the individuality and integrity of the student is truly respected. This education is based on the assumption that the student can be trusted to learn in every way which will maintain and enhance self, and he can be trusted to grow, provided the atmosphere for growth is available to him.

It has been indicated by this research that non-directive and sociodramatic techniques have contributed towards a total self-development that was reflected in increased self-insight, greater self-acceptance and a decrease in attitudes associated with racial, religious, and ethnic prejudice. Exemplary of much of this growth are several excerpts drawn from course evaluation forms submitted by students in the group therapy class:

"The most important thing that I have gained is self-confidence and a feeling that I am not the only person who feels insecure, hesitant, doubtful and quite often frustrated. I have a fuller, more rounded outlook on life. I see my friends, relatives and teachers as human beings and friends rather than scientists with microscopes waiting to study me. I feel, that, now, we are all alike. I find that I am unconsciously more sympathetic toward people and their troubles, more friendly and can certainly hold a more intelligent conversation due to the knowledge I learned this term—and it was plenty."

"As a result of this course I think I have gained a broader outlook on the 'whys' and 'hows' of the behavior patterns of people around me. It has changed my approach to everyday

living because of the technique that was used, especially those of the dramas which were used and which depicted the same everyday situations most people have gone through."

"I have gained quite a bit from this course. I have learned that most people don't do things just to be nasty but because of something that bothers them. It has taught me to have patience with my friends and relatives when they are upset and annoyed. I have further learned to control my temper."

REFERENCES

- 1) T. W. Adorno, E. Frenkel-Brunswik, D. J. Levinson, and R. W. Sanford. *The Authoritarian Personality*. New York: Harper and Brothers, 1950.
- 2) C. Rogers. *Client Centered Therapy*. Boston: Houghton Mifflin Co., 1951.
- 3) H. S. Maas. "Applying group therapy to classroom practice." *Mental Hygiene*, 35 (1951), 250-259.
- 4) H. H. Jennings. "Sociodrama as an educative process." *Fostering Mental Health in our Schools*. Washington, D. C., National Education Association, 1950.
- 5) J. L. Moreno. "Experimental sociometry and the experimental method in science," in *Current Trends in Social Psychology*, Pittsburgh: Univ. of Pittsburgh Press, 1948.

ATTITUDES OF PARENTS OF DIFFERING SOCIO-ECONOMIC STATUS TOWARD THE TEACHING PROFESSION¹

WILLIAM F. ANDERSON, JR.

University of Alabama

The purpose of the present paper is to report on an investigation of the relationships between socio-economic status and parental attitudes toward the teaching profession. The following are reported here: (1) The prestige of teaching occupations; (2) The suitability of the teaching occupations for sons and daughters; (3) The desirability of the teaching occupations for sons and daughters.

SUBJECTS AND METHODS

Expressions of attitudes were obtained by asking the parents of a selected group of high school pupils to rank three teaching occupations (grade school, high school, and college) along with a selection of other occupations representing a wide socio-economic range. The rank assigned the teaching occupation is accepted as a measure of attitude.

Measuring attitudes toward the teaching profession.—A questionnaire was constructed to facilitate the investigation of parental attitudes toward the teaching profession. In constructing this questionnaire the first step was to select two lists of occupations, one for sons and the other for daughters which (1) were common to Iowa, (2) had a comparatively large number of workers, and (3) ranged from unskilled to professional. For the part of the questionnaire dealing with prestige, ninety-three occupations were used. For the parts dealing with suitability and desirability, forty-one occupations were used for sons and twenty-nine for daughters.

Instructions to parents were in part as follows: (for prestige) "Some occupations seem more important than others. Perhaps all of us respect some occupations, and the men and women in them, more than we do others. Place a figure 1 in front of the

¹ The article is based on the writer's Doctorate dissertation, State University of Iowa, 1952. Grateful acknowledgment is expressed to Professor James B. Stroud under whose direction the work was done.

occupations you have the highest respect for. . . . Place a figure 5 in front of the occupation you have very low respect for. You may assign the same rating to as many occupations as you like."

The following are the instructions employed for suitability:

"In this part of the questionnaire we are asking you to think about each occupation and to judge how well it is suited to your son (or daughter). . . . In light of your son's (or daughter's) abilities and qualities which occupations is he (or she) best suited for? Please remember that he may be well suited to more than one occupation." Place a figure (1) in front of the occupations which are very highly suited to your son's (or daughter's) abilities and qualities. . . . Place a figure (5) in front of the occupations which are very poorly suited to your son's (or daughter's) abilities and qualities.

Parents were asked to rate the same lists of occupations (for sons and daughters) as to desirability again on a five-point scale. Here parents were asked to make the ratings in terms of how well they would like to see their son or daughter enter them.

Characteristics of the sample.—The questionnaire was administered to parents of certain of the six hundred and sixty-six tenth-grade students in the four public secondary schools of Cedar Rapids, Iowa, in the spring of 1952.

By use of the occupation of the wage earning parent, which in most cases was the father, and the address, an estimate of the social status of each pupil was made. This amounted to a preliminary screening. Each pupil was placed in one of four lists based on this initial social class estimation: upper, upper-middle, lower-middle, and lower. From each list a randomly drawn tentative sample was selected. To this sample the Index of Social Class was applied. When necessary adjustments were made, the sample appeared as follows:

	Son		Daughter		Total
	Father	Mother	Father	Mother	
Upper-class	15	15	15	15	60
Upper-middle	15	15	15	15	60
Lower-middle	15	15	15	15	60
Lower-class	15	15	15	15	60
	60	60	60	60	240

In ninety-five families both parents coöperated, while in fifty families only one parent filled out the questionnaire.

Measuring the socio-economic status.—The method selected for determining socio-economic status was the Index of Status Characteristics, I.S.C., developed by Warner, Meeker and Eells.² The formula employed for this estimation of social status is a weighted combination of values assigned to (a) Occupation, (b) Source of income, (c) House type and (d) Dwelling area.

Each parent selected in the tentative sample was personally contacted by the writer. This contacting interview had as its aim the establishing of sufficient rapport to bring about acceptance of the questionnaire. During the interview, information needed to complete the I.S.C. was gained. Where needed, corrections were made as indicated above. A date was set for personally picking up the completed questionnaires. If upon returning to the parent the questionnaire was not completed a second date was set for pick-up purposes.

Coöperation was furthered by the assistance of the superintendent of schools who wrote a letter of introduction which was shown to the parents by the interviewer. In these ways, coöperation was obtained from ninety-two per cent of the families contacted.

RESULTS

Prestige of teaching occupations.—The mean ratings on prestige for the three teaching occupations are presented in Table I. To facilitate interpretation the mean ratings of several other occupations varying widely in prestige are given. In both instances separate means are listed for the four social classes. The data are for sons only.

Noteworthy is the fact that ratings of occupations do not vary much with social class. This finding stood out in the entire list (ninety-three) of occupations, suggesting that the various occupations are held in about the same degree of esteem by parents of the four social class levels. Mickey obtained a similar result when he asked ninth-grade students to rate occupations as to prestige.³

² W. L. Warner, M. Meeker and K. Eells, *Social Class in America*, Chicago: Science Research Associates, 1949, pp. 121-159.

³ E. C. Mickey, *High School Student Rating of Occupational Prestige*, Master's thesis, State University of Iowa, 1948.

TABLE I.—MEAN PRESTIGE VALUES OF OCCUPATIONS FOR SONS BY SOCIAL CLASS OF RATERS

Occupation	Upper-class	Upper-middle Class	Lower-middle Class	Lower-class
Surgeon	1.0	1.0	1.0	1.0
College President	1.1	1.3	1.2	1.8
Lawyer	1.3	1.5	1.3	1.8
College Teacher	1.4	1.5	1.5	1.8
Supt. of Schools	1.4	2.0	1.5	1.9
High-school Teacher	1.6	1.8	1.7	1.9
Newspaper Editor	1.8	2.2	1.7	2.1
Veterinarian	1.8	2.0	2.0	1.9
Factory Manager	1.9	2.0	2.0	2.5
Rural School Teacher	2.4	2.5	2.8	2.3
Policeman	3.0	2.6	2.5	2.0
Bank Teller	2.9	2.8	2.1	2.9
Lumber Dealer	3.2	2.7	2.4	2.4
Carpenter	3.2	2.8	2.2	2.5
Hardware Dealer	3.0	2.7	2.7	2.7
Mail Carrier	3.4	2.9	2.9	2.7
Plumber	3.2	2.4	2.4	2.7
Traveling Salesman	3.0	3.0	2.8	3.8
Sales Clerk	3.6	3.4	3.7	3.6
Delivery Man	4.0	3.8	3.5	3.5
Waiter	3.8	4.0	4.1	3.9
Bartender	4.7	4.6	4.5	4.3

Also significant is the comparatively high prestige ratings accorded the teaching occupations by all social classes. This finding contrasts with the fact that important social class differences were found when parents were asked to rate occupations on suitability and desirability.

Suitability of teaching occupations for sons and daughters.—Do the views expressed by parents regarding the suitability of teaching for their sons and daughters vary with socio-economic status? The data in Table II show that the two upper classes rated all three of the teaching occupations as being more highly suited than did the two lower classes. The two lower classes rated teaching less well suited to their sons and daughters. In fact, except for lower class father's attitudes (as expressed) toward grade-school teaching for daughters it can be said that parents of the two lower classes regard teaching as poorly suited to their children. Of

TABLE II.—MEANS AND STANDARD DEVIATIONS OF THE RATINGS OF THE SUITABILITY OF TEACHING OCCUPATIONS FOR SONS AND DAUGHTERS OF DIFFERING SOCIO-ECONOMIC STATUS

		College		High School		Grade School	
		Father	Mother	Father	Mother	Father	Mother
U*: S	M & (SD)	2.7 (1.1)	3.0 (1.4)	2.3 (1.0)	2.6 (1.2)	2.7 (0.8)	3.1 (1.2)
U: D	M & (SD)	2.7 (1.1)	2.3 (1.2)	2.3 (1.1)	2.3 (1.2)	2.7 (0.9)	2.4 (1.0)
UM: S	M & (SD)	2.8 (1.4)	2.9 (1.3)	2.5 (1.2)	2.5 (1.4)	2.6 (1.3)	2.9 (1.3)
UM: D	M & (SD)	3.3 (1.2)	3.4 (1.2)	2.9 (1.1)	2.9 (0.9)	2.5 (1.0)	2.2 (0.9)
LM: S	M & (SD)	3.6 (1.3)	3.8 (1.2)	3.3 (1.5)	3.1 (1.2)	3.4 (1.5)	3.1 (1.3)
LM: D	M & (SD)	3.5 (1.5)	3.7 (1.4)	3.1 (1.3)	3.5 (1.4)	2.7 (1.0)	3.3 (1.3)
L: S	M & (SD)	3.9 (1.2)	3.9 (0.8)	3.6 (1.5)	3.8 (1.2)	3.7 (1.5)	3.9 (0.9)
L: D	M & (SD)	3.5 (1.0)	3.6 (1.4)	2.9 (1.1)	3.3 (1.5)	2.8 (1.1)	3.2 (1.3)

U* Upper class parents.

S Sons, etc.

NOTE: Due to the instructions employed in the questionnaire the lower the mean the higher the attitude of suitability.

special interest is the high rating given high-school teaching by the two upper classes, the second and fourth highest places among the forty-one occupations. In the two lower classes, high-school teaching was ranked nineteenth and thirty-first with grade-school and college teaching ranking even lower.

Mothers of the several classes differed from fathers in that they judged all three teaching occupations a little less suitable for their sons—with the exception of the mothers of the lower-middle class.

As a means of determining the statistical significance of social class differences in ratings, and differences among the ratings of the three occupations, analyses of variance (f test) were computed. Examination of these data on the fathers' ratings revealed that the differences among the social classes concerning the suitability of the three teaching occupations for their sons were statistically significant at the five per cent level.⁴ The analysis of variance showed that differences in the ratings of the three teaching occupations were also significant.

Mothers of the four social classes differed significantly (f test) in rating the suitability of the teaching occupations for their sons. This analysis also revealed that mothers did not judge the three teaching occupations as equally well suited for sons but rated

⁴ Hereafter, when significant difference is used it refers to a difference that is statistically significant at the .05 level or better.

high-school teaching as significantly better suited than either grade-school or college teaching.

Parents with daughters judged the suitability of the three teaching occupations differently than did parents with sons. As can be seen in Table II, college teaching is strictly an upper class occupation; only this social class rated it as well suited for daughters. A much different attitude exists about the suitability of grade-school teaching; fathers of all four social classes rated this teaching occupation above average while mothers of the more favored environments rated it highly suitable and mothers of the two lower classes rated it slightly below average.

With respect to high-school teaching, parents of the upper class rated this teaching occupation highly suitable for daughters while the upper-middle class rated it somewhat lower but still above average. Fathers of the two lower classes rated it average in suitability, while mothers judged it somewhat below average.

Once again mothers differed from fathers in judging the suitability of the three teaching occupations. Ratings of mothers of the upper class were higher than those of fathers, but in the less favored environments, the opposite trend was noticed.

The analysis of variance of the ratings of fathers with daughters revealed statistically significant differences in the way the three teaching occupations were rated, but non-significant social class differences. Analysis of variance of mothers' ratings revealed both statistically significant social class and occupational differences.

Desirability of teaching.—Parents of different socio-economic status were asked to think less realistically and to judge just how much they would 'like' to see their son or daughter enter the occupations previously rated as to suitability.

Means and standard deviations of ratings of the teaching occupations as to desirability are found in Table III. Since the differences between the ratings of mothers and fathers were so small, the attitudes of parents have been combined.

The only difference that is significant among the ratings of the upper and lower social classes on desirability is that for the category—college teaching. Among the forty-one occupations rated on desirability for sons college-teaching was ranked the fifth and third most desirable by the upper and upper-middle classes, respectively. High-school teaching was rated ninth and tenth

TABLE III.—MEANS AND STANDARD DEVIATIONS OF THE DESIRABILITY OF TEACHING OCCUPATIONS FOR SONS AND DAUGHTERS OF DIFFERING SOCIO-ECONOMIC STATUS

		College		High School		Grade School	
		Son	Daughter	Son	Daughter	Son	Daughter
Upper-class	Mean	1.8	1.7	2.5	2.0	2.9	2.8
	SD	1.0	0.7	0.9	1.1	1.1	1.2
Upper-middle	Mean	1.8	2.5	2.3	2.4	2.8	2.2
	SD	0.8	1.0	1.2	1.0	1.1	1.0
Lower-middle	Mean	2.4	2.7	2.8	2.9	3.2	2.7
	SD	1.5	1.5	1.4	1.4	1.4	1.3
Lower-class	Mean	3.1	2.7	2.8	2.2	3.1	2.3
	SD	1.5	1.3	1.4	1.3	1.4	1.0

NOTE: The lower the mean the higher the attitude of desirability.

most desirable by these same classes and was judged more desirable than such professional occupations as newspaper reporter, draftsman, insurance agent and pharmacist. Although rated somewhat lower, grade-school teaching was judged more desirable by these two classes than all the skilled occupations. This was also true in the lower-middle class with the exception of electrician, carpenter and machinist, which were judged more desirable than grade-school teaching. The upper-class judged college-teaching to be the most desirable for daughters of all twenty-nine occupations rated, with high-school teaching judged more desirable than such professional occupations as office manager, registered nurse, social worker, stenographer, purchasing agent and newspaper reporter. The upper-middle class judged grade-school teaching to be the second most desirable for daughters of all occupations rated, with high-school and college teaching rated only slightly lower. The two lower classes rated teaching only slightly less desirable for daughters than did the favored classes, although they rated certain business occupations (stenographer, typist and bookkeeper) as being slightly more desirable than teaching. Other business occupations were rated considerably less desirable than teaching.

It is an interesting observation that although teachers tend to

come from the lower and lower-middle classes⁵ (probably making them about average proportionately with respect to social class), parents of these two classes, as suggested by this investigation, appear to regard teaching as a less suitable and less desirable occupation for their sons and daughters than do parents in the two upper classes.

CONCLUSIONS

1) In rating occupations as to prestige value the parents of this sample showed little difference one social class from another. Mean ratings for the four social classes were remarkably similar for the four classes. All three of the teaching occupations dealt with in this paper ranked relatively high on the prestige scale. Among the ninety-three occupations the composite rank value on a five-point scale for college teaching was 1.6 (1.1 being highest); for high-school teaching, 1.8; and for grade-school teaching, 1.9. (Eight occupations ranked higher than college teaching, in this order: surgeon, physician, college president, judge, clergyman, architect, dentist and electrical engineer.)

2) Ratings on suitability of the teaching occupations varied with social class status. Parents of the upper two classes rated the teaching profession more suitable for their sons and daughters than did parents of the lower two classes.

3) Ratings on desirability of teaching as an occupation also varied with social class status, although to a lesser degree, except for college teaching.

4) Ratings on the suitability of the three teaching occupations varied as to the sex of the pupil. For all social classes combined, parents judged high-school teaching to be significantly more suitable for sons than the other two teaching occupations. The upper three classes rated college teaching as being more desirable for their sons. With the exception of the upper class, grade-school teaching was judged to be better suited to daughters than the other two teaching occupations. It was also rated as being much more desirable for daughters than for sons.

⁵ Nathan A. Edwards, *Sociology of Teaching. II: A Study of the Male Classroom Teacher*, Doctor's dissertation, State University of Iowa, 1952.

COMPARATIVE DECLINE OF GRADUATE RECORD EXAMINATION SCORES AND INTELLIGENCE WITH AGE

R. TRAVIS OSBORNE and WILMA B. SANDERS

University of Georgia

From the literature (3, 4, 7, and 8) it is possible to infer that adult age increments are accompanied by a decline in acquired knowledge which would parallel the decline in intelligence with age. However, with the exception of Learned and Wood's *The Student and His Knowledge* (3) and a previous study by the authors (4) there is almost a complete lack of direct data on the comparative decline of acquired knowledge and intelligence of adults. This investigation is for the purpose of securing research evidence relevant to this problem.

POPULATION AND PROCEDURE

In the present study the Graduate Record Examination was used for determining the student's level of acquired knowledge. This test according to the G.R.E. prospectus "is the only examination of national character which is expressly designed to measure achievement in the principal fields of collegiate study at the advanced level and which is endorsed or required by graduate and professional schools" (2). Actually the G.R.E. is a series of subject-matter tests designed to cover broadly the main areas of the typical undergraduate program of instruction: Mathematics, Physics, Chemistry, Biological Science, Social Science, Fine Arts, Literature, and Verbal Factor. The latter test is not a content test in the true sense of the word but it was developed and included in the G.R.E. battery to measure the ability to discriminate word meanings.

The one thousand eight hundred and seven graduate students who were in attendance at the University of Georgia between February, 1946, and February, 1952, were subjects for study. This group includes all graduate students except some few irregular students who were not seeking an advanced degree. The seven professional schools and the five divisions of the College of Arts and Sciences were represented. Men students outnumbered

women: one thousand one hundred and seventy-three to six hundred and thirty-four.

G.R.E. scores, personal data, and college marks were available from the files of the Dean of the Graduate School. Machine methods were used for processing and conventional statistical procedures were used in analyzing the data.

Figures 1 and 2 compare graphically G.R.E. age-score and intelligence curves. The two sets of data were not obtained from the same subjects; the age-achievement curve was plotted from the results of G.R.E. and the intelligence curve from Wechsler's normative data. For comparative purposes these curves have been reduced to approximately the same scale.

The number, mean, and standard deviation for the seven profiles and three G.R.E. averages were determined for each four-year age increment from nineteen to fifty-five (Table I). Also, not shown but available on request there are tables of inter age-group confidence levels of mean differences for the seven achievement profiles.

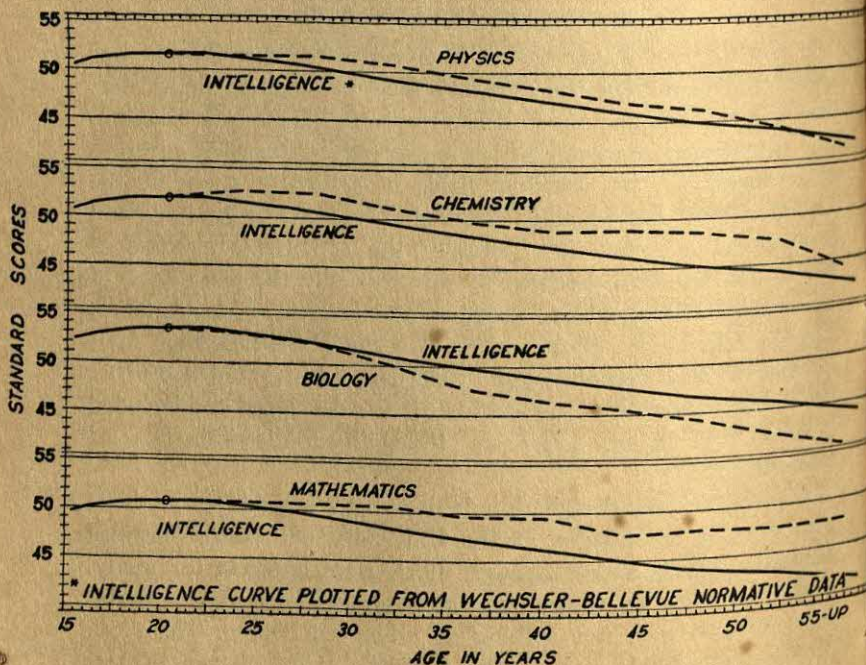


Fig. 2—Variations in G.R.E. Science Profiles and Intelligence with Age.

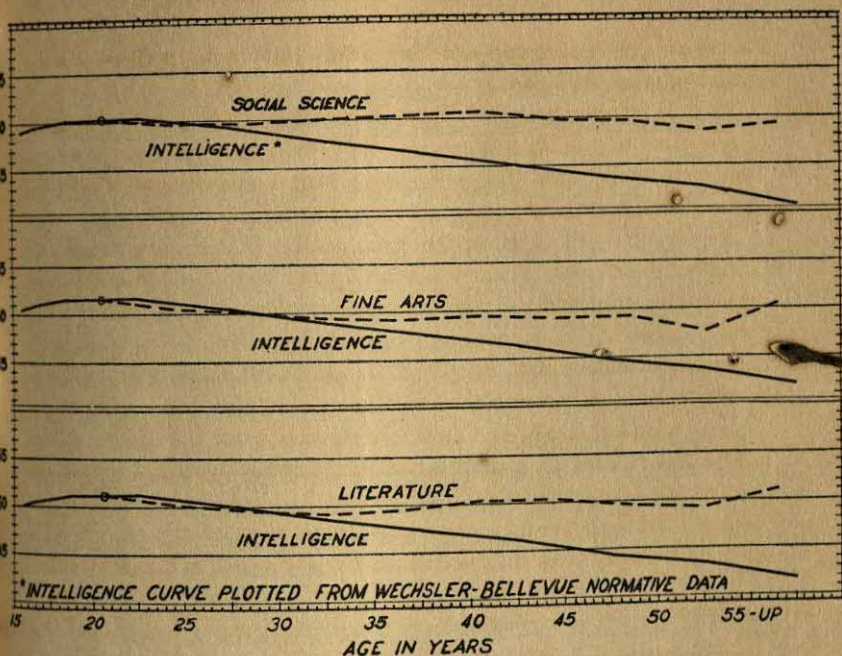


FIG. 2.—Variations in G.R.E. Non-Science Profiles and Intelligence with Age.

RESULTS

Examination of Figure 1 shows that for the Sciences adult achievement loss with age is strikingly similar to the decline of mental ability with age. For Physics and Chemistry the age-score curves almost parallel the declining mental ability curve. Maximum performance in Chemistry is maintained slightly longer and the decline is somewhat more gradual than for Physics. Mean scores in Biology, however, fall off sharply after reaching a peak in the nineteen to twenty-two age level. The loss here is greater than the corresponding loss in intelligence with age. Among the sciences, mathematics scores seem to hold up best. Significant losses were not noted until after age thirty-eight. This loss is considerably less than the corresponding loss for these ages in intelligence.

For Social Science there is no significant falling off of mean achievement scores with age. The greatest mean difference is only 3.6, and that is found between two adjacent age groups in

TABLE I

GRADUATE RECORD EXAMINATION PROFILE TEST SCORES FOR
UNIVERSITY OF GEORGIA GRADUATE STUDENTS, BY AGE

AGE GROUP	PHYSICS				CHEMISTRY				BIOLOGICAL SCIENCES				MATHEMATICS			
	No.	MEAN	ST.	S.D.	No.	MEAN	ST.	S.D.	No.	MEAN	ST.	S.D.	No.	MEAN	ST.	S.D.
19-22	132	50.39	9.68		128	51.40	10.87		132	54.54	9.14		91	49.27	9.35	
23-26	413	51.86	9.19		400	52.05	10.31		413	53.26	9.14		339	51.10	10.16	
27-30	354	51.33	10.21		326	50.78	9.94		354	52.00	9.20		288	50.79	7.42	
31-34	216	50.84	9.41		202	50.25	9.50		216	49.59	9.71		164	49.43	10.81	
35-38	205	49.52	10.00		175	47.84	9.07		204	46.92	9.60		153	51.23	11.45	
39-42	155	46.92	9.68		142	47.66	8.94		155	45.13	8.74		118	47.71	9.82	
43-46	98	47.42	10.81		88	47.71	9.50		98	45.79	10.57		68	47.49	11.81	
47-50	54	46.50	9.19		53	48.07	10.00		54	43.62	9.02		41	47.40	10.33	
51-54	34	42.01	8.66		33	44.07	8.32		34	40.39	9.25		25	49.95	9.84	
55 UP	22	41.59	9.41		21	44.21	5.65		22	41.26	10.00		14	38.64	12.04	

AGE GROUP	FINE ARTS				LITERATURE				SOCIAL SCIENCE				G.R.E. AVERAGES		
	No.	MEAN	ST.	S.D.	No.	MEAN	ST.	S.D.	No.	MEAN	ST.	S.D.	PHY.	Soc.Sc.	TOTAL
19-22	132	52.96	10.65		132	52.82	9.73		132	50.55	9.74		52.12	52.11	51.82
23-26	410	50.97	9.95		413	50.49	9.89		413	50.08	9.43		52.39	50.51	51.47
27-30	343	49.29	10.00		352	48.82	10.27		354	49.28	10.10		51.39	49.13	50.31
31-34	211	49.31	10.30		216	49.07	10.33		216	50.11	10.26		50.23	49.50	49.81
35-38	202	49.21	9.90		204	49.66	10.05		203	51.02	11.19		48.11	49.96	49.30
39-42	152	48.93	9.55		155	50.38	9.12		155	49.67	9.74		46.54	49.66	48.07
43-46	96	51.73	9.45		98	51.62	10.44		97	51.38	10.41		46.95	51.58	49.13
47-50	53	44.88	7.84		54	47.91	8.08		54	47.74	8.13		46.28	46.86	46.68
51-54	33	48.92	8.69		34	47.56	8.90		34	48.34	8.45		42.14	48.27	45.73
55 UP	22	52.43	9.40		22	54.27	9.69		22	50.51	9.59		42.32	52.40	46.58

*UNIVERSITY OF GEORGIA NORMS (N=1716)

the forties, suggesting the probability of sampling errors or factors other than significant age-score trends (Table I). The smoothed Social Science age-score curve does not follow the mental ability curve but forms a line almost parallel to the base through the group mean (Figure 2).

Fine Arts and Literature means not only hold up with age but also some statistically significant mean differences favor the older student (Table I). Only from about twenty to twenty-five do these smoothed age-score curves parallel the intelligence curve (Figure 2). As was the case for Social Science, the Literature and Fine Arts age-score curves approximate lines parallel to the base through the group mean. There is almost an imperceptible tendency for these three tests to become elevated in the forty to forty-five age range, level off, and then rise again in the over fifty-five group.

DISCUSSION AND SUMMARY

The primary purpose of this study was to compare the decline with age of acquired knowledge and intelligence. Previous investigations have implied that there could be expected a gradual decline in achievement test scores after about age twenty-four to twenty-five, the age according to Wechsler at which "our intellectual capacities are definitely on the decline" (8). However, in so far as the present writers have been able to determine, no previous attempt has been made to compare the rates of decline in adult achievement and intelligence with age. Peterson did report in 1943 that "for each age there is a trend toward a lowered G.R.E. average score with increasing age" (6), but his findings were based on only thirty-two cases in the over twenty-nine age group and he made no attempt to compare intelligence and achievement decrements. By averaging G.R.E. profile scores the differential rates of decline were lost and the slope of the age-score curve was attenuated.

There are several possible interpretations of the results presented above. For example, Peterson attributes the decline in G.R.E. scores with age to "the fact one loses considerable facility or becomes 'rusty' in a subject for which one has little or no opportunity for practice while one loses little in a function which is constantly practiced" (6). On the other hand if the G.R.E. is just a long test of verbal intelligence (5) it would naturally be expected to show the characteristic decline of mental ability with age.

Although research evidence is lacking on the point, Gilbert believes "motivation plays an important part in intellectual decline with advancing age, for when motivation is absent, we find waning interest in things and a decline in intellectual activity and use of intellectual abilities. The result of this disuse, of course, is accelerated decline" (1). Still other explanations may occur to the reader; such as, for example, the inadequate sampling techniques inherent in cross-sectional studies. The fact that if the older age groups are found to be heavily weighted with women and/or Social Science, Literature, or Fine Arts majors would account for the sustained mean scores in these areas.

The results presented in the previous section would appear to support the hypothesis of a differential rather than a consistent

rate of decline of acquired knowledge with age. Only the science profiles of the G.R.E. follow the characteristic curve of mental decline with age. Social Sciences, Fine Arts, and Literature scores seem to hold up well even at the extremes of the age range. As the rates of decline are largely functions of the nature and content of the tests used these findings should not necessarily be expected to hold for different tests, other age groups, or for any individual student.

REFERENCES

- 1) Jeanne G. Gilbert. *Understanding Old Age*. New York: The Ronald Press Co., 1952. 422 pp.
- 2) *The Graduate Record Examination and Returning Service Personnel*. A Descriptive Prospectus for Men and Women in the Armed Services Who Desire Advanced Collegiate and Graduate Study After the War. Boston: The Merrymount Press. 11 pp.
- 3) William S. Learned and Ben D. Wood. *The Student and His Knowledge*. Bulletin #29. New York: The Carnegie Foundation for the Advancement of Teaching, 1938. 406 pp.
- 4) R. Travis Osborne and Wilma B. Sanders. *Differential Decline in Graduate Record Examination Scores with Age*. Unpublished paper read before the meeting of the Georgia Academy of Science, Macon, Georgia, April, 1953.
- 5) Katherine L. Pease. *Organization of Ability on the College Level: A Study of Graduate Record Examination Profile Tests*. Unpublished Doctor's dissertation, Columbia University, New York, 1949. 84 pp.
- 6) Stuart Conrad Peterson. *The Measurement and Prediction of Scholastic Achievement on the Graduate Level*. Unpublished Doctor's dissertation, State University of Iowa, Iowa City, 1943. 121 pp.
- 7) David Wechsler. *The Measurement of Adult Intelligence*. Baltimore: The Williams & Wilkins Co., 1944. 258 pp.
- 8) ——. *Range of Human Capacities*. Baltimore: The Williams & Wilkins Co., 1952. 190 pp.

THE SOPHOMORE GENERAL CULTURE TESTS AND STUDENT KNOWLEDGE OF CULTURAL FACTS

BERT D. ANDERSON

Southern State College, Magnolia, Ark.

The problem of appraising the college student's knowledge of cultural facts has received a good deal of attention in recent years. Special interest in this matter is justified since educators generally agree that American colleges have the responsibility of acquainting students with scientific and cultural information that is felt to be basic in perpetuating our society. Various programs of general education have been devised to improve the student's knowledge of his heritage and the responsibilities he must accept if he is to become a worthy citizen in these critical times. In all of these programs it has been assumed that if the student possesses a certain body of general information he is more likely to become an effective citizen in our democracy.

Perhaps the Sophomore General Culture Test in its various forms has been more widely used than any other single instrument to estimate the extent to which the college student has acquired a knowledge of cultural facts that are assumed to be significant in our society.¹ It is common practice to administer this test at the end of the sophomore year when the student has completed most of his general education training.

The major research pertaining to this instrument has been conducted on a national basis of the Educational Testing Service through its national college testing program. Their findings have enabled individual colleges to compare the performance of their students on this test with that of a national college group. They have demonstrated that there is a wide variation in scores made within and between colleges on the Sophomore General Culture Test.^{2, 3, 4, 5} However, for the local administrator of the general

¹ One hundred seventy colleges representing thirty-seven states participated in the 1950 College Sophomore Testing Program.

² Committee on Educational Testing, "Report of the sophomore testing program," *Educational Record*, xiv, 1933, pp. 522-71.

³ The 1932 College Sophomore Testing Program, "A report by the Ad-

education program and the counseling staff these very practical questions arise: (1) How comprehensive is the coverage of this test when its items are related to course offerings found within the general education program? (2) How should the student's performance on this test be interpreted when related to the instruction he has received in the local general education program? (3) What do these scores tell us about the student's knowledge of cultural facts?

A partial answer to the above questions was found at the University of Utah by (a) tabulating the Sophomore General Culture Test items according to the courses to which they were related that were offered within the general education curriculum and (b) ascertaining the relationship between scores received on the General Culture Test and the student's past achievement in four separate areas of general education.

SAMPLE

In the spring of 1950 the Sophomore General Culture Test, Form Y, was administered to eight hundred and forty-nine students who were completing their sixth quarter of work at the University of Utah. They had finished most of their general education classwork requirements in the Physical Sciences, the Biological Sciences, the Humanities, and the Social Sciences. Two hundred students, whose scores and stated professional aims were found to be representative of the total group, were selected for more intensive study and used as the sample for this research.

FINDINGS

I) *Test items coverage of curriculum content.*—Students were required to complete at least nine credit hours in each of four general education areas. These were the Biological Sciences, the Physical Sciences, the Social Sciences and the Humanities. The number of courses made available to the student varied from ten in the Biological Sciences and the Physical Sciences to thirty-

visory Committee on College Testing," *Educational Record*, XIII 1932, pp. 290-343.

⁴ "The Report of the 1936 College Sophomore Testing Program," *Educational Record*, XVII, 1936, pp. 577-608.

⁵ "Report on the 1950 National College Sophomore Testing Program," Coöperative Test Division, *Educational Testing Service*. pp. 1-4.

TABLE I.—NUMBER OF TEST ITEMS DEVOTED TO SUBJECT MATTER TAUGHT IN THE SOCIAL SCIENCE AREA OF GENERAL EDUCATION

General Education Area and Name of Course	Number of Courses Offered	Related Current Soc. Prob. Test Items	Related Hist., Soc. Studies Items
Social Science Area			
Economics	6	1	7
Geography	2	0	1
History	11	1	48
Health*		8	0
Anthropology	1	0	0
Political Science	7	19	14
Sociology	3	25	0
Psychology	2	0	4
Philosophy	2	0	0
Social Sc. Survey	3	U.C.**	U.C.**
	37	54	74

* Not offered in this general education area.

** Unclassified.

seven in the Social Science areas. The various courses in each of the four areas were listed separately so that the items of the appropriate Sophomore General Culture Test could be related to the course content being offered. Some of these findings are presented in Table I.

In general, the test items found within the Current Social Problems and History and Social Studies Tests provided inadequate coverage of course subject matter being offered in the Social Science area. While there were forty-nine History questions contained in these two tests, it should also be noted that there were eleven separate History courses offered. There were no questions pertaining directly to Philosophy or Anthropology, yet these two courses were regularly being taken by students fulfilling the general education requirement in this area.

In the other three areas, as was the case in the Social Sciences, course offerings were in many instances covered by too few items or none at all. Only ten courses were offered in the Biological Science area as compared to thirty-seven in the Social Sciences; however, only twenty-seven items of the Science Test were directly related to Biological Science courses while the remaining thirty-

three items were designed to cover subject matter taught in ten courses offered in the Physical Science area. In specific subject matter situations the lack of coverage was also noted. For instance, the Literature and Fine Arts Tests, each containing ninety items, could best be applied to the courses offered in the Humanities area. A review of these items revealed that there were no questions asked in either test about such famous philosophers as Berkely, James, Bergson, Russell or Nietzsche, yet the Philosophy offerings within the Humanities area stressed the works of these men. The General Culture Mathematics Test, while perhaps important from other standpoints, bore little or no direct relationship to subject matter offered in any of the four general education areas.

II) *Achievement in General Education related to scores received on the General Culture Test.*—To discover the relationship between the student's past achievement in courses completed in the Biological Sciences, the Physical Sciences, the Humanities, and the Social Sciences, separate grade point averages were computed for each area and used as the criterion of achievement. These grade point averages were then correlated with the scores he had received on the General Culture subtests. The Pearson Product Moment correlation method was used to establish these relationships. The results are reported in Table II.

Each subtest, as well as the total scores of the Sophomore General Culture Test, correlated positively with achievement in all

TABLE II.—THE CORRELATION COEFFICIENTS FOUND BETWEEN THE GRADE-POINT AVERAGES IN EACH OF THE GENERAL EDUCATION AREAS AND SCORES RECEIVED ON THE SOPHOMORE GENERAL CULTURE SUBTESTS

General Education Area	Current Social Prob.	History Social Studies	Lit.	Science	Fine Arts	Math.	Tot. Gen. Cult.
Biological Science Grade Point Average	.374	.366	.311	.507	.320	.357	.506
Physical Science Grade Point Average	.211	.315	.288	.526	.147	.429	.422
Social Science Grade Point Average	.344	.333	.374	.356	.289	.254	.444
Humanities Grade Point Average	.146	.204	.282	.209	.262	.076	.281

four general education areas. With few exceptions, these relationships were found to be significant at either the one or five per cent level of confidence. Only the Science Test and the Total General Culture scores correlated well with the grades earned in the Physical Sciences and the Biological Sciences. The total score was found to be the best indicator of the student's past achievement in the Social Sciences while none of the General Culture subtest scores were closely related to achievement in the Humanities. The total score was the most consistent index of past achievement in general education even though there was a rather wide range in the size of these correlation coefficients.

DISCUSSION

The breadth of subject matter covered within the general education program presents an almost overwhelming problem when an attempt is made to measure student knowledge in this area. With so many courses being offered in each of the general education areas, it is only reasonable to suspect that many different course combinations will be presented by students in fulfilling the general education requirements. The difficulties involved in constructing an instrument that will adequately cover all of these subject matter areas becomes clearly recognized. With only a few items to cover several complete courses in some cases, perhaps it was not at all surprising to find low correlation coefficients between the grade-point averages earned in each of the four areas and the student's test performance.

The lack of relationship between test scores and achievement and the very limited coverage of test items leads one to suspect that the student's General Culture scores will not necessarily indicate how effective the local general education instruction has been in improving the student's knowledge of cultural facts. On the other hand, neither does the grade-point average earned in each of these four areas give a well-rounded picture of student knowledge of cultural facts, since it is based upon a limited performance in two or three courses in a broad subject matter area. For instance, the student might complete the Humanities requirement without ever having completed a course in Literature. Perhaps it is at this point that the General Culture Literature Test can serve a useful purpose. If, in such a case, the counselor is willing to assume that the student's performance on the Litera-

ture test represents a rather independent but objective index of the student's knowledge of literature, then these scores can be used to supplement grades in appraising the student's knowledge of cultural facts in the Humanities area. If these General Culture Test results are regarded more or less as an independent index of cultural knowledge then their validity must be defended upon the premise that the test is not only technically sound, but is based upon subject matter that is educationally significant. The lack of relationship between General Culture test scores and past achievement in General Education courses does tend to support the hypothesis that these results do provide an independent measure of student knowledge in these general education areas. It seems quite clear that no single measure will give a well-rounded picture of the student's knowledge of cultural facts.

CONCLUSIONS

- 1) The General Culture test does not provide adequate test item coverage when a number of elective courses are made available to students completing their general education requirements.
- 2) The General Culture test results do not indicate how effective the local general education instruction has been in improving the student's knowledge of cultural facts.
- 3) The General Culture test results may be used as an independent index of the student's knowledge of cultural facts if it can be assumed that the test is technically sound and based upon subject matter that is educationally significant.

FINAL EXAMINATION PERFORMANCE IN TRADITIONAL AND CROSS-DISCIPLINARY COURSES

IRWIN A. BERG and RAYMOND J. MURPHY¹

Northwestern University

While the number of interdepartmental or cross-disciplinary courses appears to be increasing among American colleges and universities (2, 3), the rise in popularity of such courses has been accompanied by pointed questions as to their academic value when compared to the traditional, single-subject courses. For beginning college students, the traditional course is regarded by many as more 'solid' and as providing a firmer basis for mastery of the subject matter, while the cross-disciplinary courses are thought to encourage superficial and popularized treatment of several fields. This attitude, understandably, has been indignantly challenged by proponents of inter-departmental course offerings. Curiously, nothing seems to have been published which offers factual data bearing upon the problem.

In 1949, with the support of a grant from the Carnegie Foundation, a cross-disciplinary course known as An Introduction to the Sciences of Human Behavior was jointly offered by the Departments of Anthropology, Psychology, and Sociology of Northwestern University. The course was not intended to be merely a survey of these fields. Rather, the aim was explicitly one of providing the students enrolled in the course with the same concepts, the same fund of information had they taken the three introductory courses in anthropology, psychology, and sociology. Thus the inter-disciplinary course was intended to serve as a prerequisite for advanced work in any or all of the three fields in the same way as the traditional courses. But more than that, by bringing the fruits of scientific research and the explanatory concepts of the three disciplines to bear upon a general aspect of human behavior, it was expected that the students who completed the interdisciplinary course would achieve a broader and firmer grasp of the subject matter than stu-

¹ The authors wish to express their thanks to Professor Kimball Young for his critical reading of the manuscript.

dents who took the three traditional introductory courses separately.

The present study attempts an answer to the question of whether students who learned their elementary psychology in the cross-disciplinary Sciences of Human Behavior have mastered the subject matter as well as those who completed the regular or traditional beginning psychology courses. While the human behavior students were exclusively freshmen and the psychology content of their course was spread over three academic quarters, it was nevertheless believed that a reasonable comparison could be made of the two groups even though the traditional elementary psychology course students were sophomores and above and had their psychology in a one-quarter course. The basis for comparison was sixty multiple-choice test items which were administered to both the human behavior and traditional psychology classes by including these items in the final examinations during the academic year 1951-52. Twenty items were administered in each of the fall, winter, and spring quarters to a total of three hundred and eighty-four traditional psychology course students and to approximately three hundred human behavior students. Although the sixty test items sampled the entire range of subject matter for elementary psychology, only twenty items representing about one-third of the total material could be given each quarter. This division was necessary because the human behavior students have the psychology content of the course distributed over the entire academic year, while the traditional psychology students complete their course in one quarter. The total hours spent in class and the textbook assignments were the same for both groups. The test items were selected from the reading assignments required of both groups. Agreement among the human behavior and regular psychology instructors was used in choosing the items for inclusion in the final examinations. The test items which were identical for both groups were placed in a block at the beginning of each examination. The mean score for each examination was computed for both groups and for each academic quarter, and the significance of the mean score difference tested by determining critical ratios.

Because a variety of entrance tests are used by the University's Office of Admissions, it was possible to secure scholastic aptitude test scores for only one hundred and eighteen human behavior students and forty-one traditional psychology students who had

taken the same form of *Northwestern Analogies Test* (1). A check of the general academic performance of these students for whom Analogies scores were available indicated that the groups could be regarded as roughly equivalent. Comparisons were also made of the identical item scores for students from the various colleges of the University. This was regarded as necessary because the human behavior course draws its enrollment chiefly from the College of Liberal Arts while the traditional psychology course includes students from virtually every school on the campus. Since the proportion of male to female students in both courses was about the same, no control over this factor was deemed necessary.

DISCUSSION

Viewed as a pattern for the year, the performance of the human behavior and regular psychology groups may be regarded as reasonably comparable. As reference to Table 1 will show, the traditional psychology group earned a significantly higher mean score on the final examination items for the fall quarter; however, the human behavior group performance was slightly higher, though not significantly, on those items given to both groups for the winter and spring quarter examinations.

Two factors probably account for much of the difference in the mean examination scores of the two groups for the fall quarter. First, the course plan of the human behavior group requires that a considerable portion of the fall quarter be devoted to an elaboration of the terms and concepts used in anthropology, sociology, and psychology. While this time is well spent, since it provides a solid foundation for the rest of the year's work, it does mean that the

TABLE 1.—FINAL EXAMINATION PERFORMANCE OF TRADITIONAL AND HUMAN BEHAVIOR GROUPS

Course	Fall Quarter				Winter Quarter				Spring Quarter			
	N*	M	SD	CR	N*	M	SD	CR	N*	M	SD	CR
Traditional	150	13.3	2.8	3.9	114	11.9	3.2	.9	120	13.9	2.6	.5
Human Behavior	300	12.2	2.9		265	12.2	2.5		260	14.1	2.6	

* The number of traditional psychology students represents different classes during each quarter. The human behavior group varies in number because of absences from the final examination.

cross-disciplinary students have a temporarily leaner academic fare for the psychology portion of the fall quarter when compared to the regular psychology course. The second and perhaps more important factor is that of academic experience. Being younger and completely new to the college situation, the human behavior students should be most handicapped during the fall quarter when compared to the older and academically more experienced traditional psychology students. This advantage should be reduced in the winter and spring quarters as the human behavior group improved in study skills, became familiar with multiple-choice tests, etc. Such was actually the case as shown in Table 1.

Thus far our data indicate that the traditional psychology and the human behavior groups were, from an over-all standpoint, performing about equally on the sixty psychology examination items administered during the year. However, while total final examination performance was approximately the same for both groups, it is emphasized that the interdisciplinary students were performing as freshmen about as well as the regular psychology students were performing as sophomores or above. While the two groups were quite comparable in over-all performance on the psychology test items, it cannot be argued that the human behavior students would necessarily have benefited even more had they been sophomores, as were the regular psychology students. True, the interdisciplinary students, though freshmen, demonstrated a level of mastery of psychology which was comparable to sophomores. However, it has not been shown that the regular psychology students would not have done just as well had they taken the course in the freshman, instead of the sophomore, year. Furthermore, although many of the weaker students leave college during the freshman year, it is possible that motivation for topflight performance may decrease among the more capable students in the sophomore year, once they have tested their academic mettle as freshmen and found it satisfactory. The freshmen in the human behavior course learned as much psychology as the sophomores in the traditional psychology course, according to the data presented; and that is about all that can be said in the light of available information.

A question which naturally arises is whether one group was brighter than the other. Since scholastic aptitude test scores were available for only a limited number in both groups, no very positive statement can be made about the influence of this factor. As shown

TABLE 2.—NORTHWESTERN ANALOGIES TEST SCORES FOR SPRING QUARTER 1952

Course	N	M	SD	CR
Traditional	41	87.1	20.6	1.0
Human Behavior	118	91.3	22.2	

in Table 2, the evidence we have indicates that there was no significant difference between the two groups. The grade point average for all courses of the forty-one regular psychology students was almost exactly at the mean of the grade point average for the total sophomore class. The mean Analogies test score for these forty-one students was 87.1, while the mean Analogies score for all students who took the test at the same time was 86.7. Thus it appears that the regular psychology group sample was probably a representative one in terms of performance on the test and for grades earned. The one hundred and eighteen interdisciplinary freshmen for whom Analogies scores were available earned a grade point average which was somewhat above that of the freshman class. The mean Analogies score for these one hundred and eighteen freshmen was 91.3, while a mean score of 86.4 was earned by all entering freshmen. It is suggested that these data strengthen the belief that the human behavior and traditional psychology students were comparable in scholastic aptitude since the slight, but non-significant, difference in Analogies test performance seems very likely to be due to a minor sampling bias. That is, the sample of human behavior students had a slightly higher mean scholastic aptitude score but they also earned slightly higher grades than the sample of the regular psychology students. This is what would be expected if the sample was slightly biased in favor of the human behavior group.

As noted earlier, the regular psychology course draws its enrollment from nearly every college of the University while the interdisciplinary course draws chiefly from the College of Liberal Arts. If the students from one college were notably stronger or weaker academically, the mean final examination scores might be influenced by such a deviant group. Tables 3, 4, and 5 present data bearing upon this point. While the mean psychology final exami-

TABLE 3.—FINAL EXAMINATION PERFORMANCE OF TRADITIONAL PSYCHOLOGY GROUP ANALYZED BY SCHOOLS FOR SPRING QUARTER 1952

School	N	Per cent	M	SD
Commerce	43	35.8	13.7	2.3
Liberal Arts	20	16.7	14.7	2.5
Engineering	18	15.0	13.2	2.9
Music	11	9.2	15.5	3.4
Education	10	8.3	13.3	2.3
Speech	9	7.5	13.5	1.6
Journalism	9	7.5	14.8	2.2
Total	120	100.0	13.9	2.6

TABLE 4.—CRITICAL RATIO MATRIX OF FINAL EXAMINATION MEAN SCORE DIFFERENCES AMONG SCHOOLS REPRESENTED IN TRADITIONAL PSYCHOLOGY FOR SPRING QUARTER 1952

School	Liberal Arts	Engineering	Music	Education	Speech	Journalism
Commerce	1.4	.8	1.9	.5	.3	1.3
Liberal Arts		1.6	.2	1.3	1.2	.1
Engineering			1.6	.1	.3	1.5
Music				1.6	1.6	.5
Education					.2	1.4
Speech						.9

TABLE 5.—FINAL EXAMINATION MEAN SCORE DIFFERENCES BY SCHOOLS REPRESENTED IN THE HUMAN BEHAVIOR COURSE FOR SPRING QUARTER 1952

School	N	Per cent	M	SD
Liberal Arts	192	73.8	14.0	2.6
Education	48	18.5	14.3	2.8
Other	20	7.7	14.1	2.9
Total	260	100.0	14.1	2.6

nation scores varied somewhat for students from the different colleges, none of these differences were of statistical significance at even the five per cent level of confidence. Thus there is no evidence to indicate that differential performance on the psychology final

examination items had any significant influence on the size of the total group means.

In conclusion, it appears reasonable to state that our evidence shows that the freshmen in the interdisciplinary course learned as much psychology as the sophomores and above in the traditional course. Whether the interdisciplinary students would have demonstrated even greater achievement had they been sophomores, or whether the traditional psychology students would have performed as well or better had they taken elementary psychology as freshmen, is not known. No evidence is available which bears on this point.

SUMMARY

A study was made of the final examination performance of three hundred and eighty-four students in the regular or traditional psychology course and approximately three hundred students in the interdisciplinary human behavior course which is offered jointly by the departments of anthropology, psychology, and sociology at Northwestern University. Sixty test items which sampled the range of material covered in elementary psychology were administered to both groups as part of the final examinations during 1951-52. Though composed entirely of freshmen, the interdisciplinary students in over-all performance did about as well as the traditional course students who were sophomores and above. No evidence was found that the two groups differed significantly in scholastic aptitude.

REFERENCES

- 1) E. L. Clark. *The Northwestern analogies test*. Evanston, Illinois: 1949.
- 2) H. T. Morse (Ed.). *General Education in Transition: A Look Ahead*. Minneapolis: University of Minnesota Press, 1951.
- 3) I. T. Sanders. "Societies around the world." *J. Educ. Sociol.*, 1947, 21, 238-242.

BOOK REVIEWS

HELEN M. WALKER AND JOSEPH LEV, *Statistical Inference*. New York: Henry Holt and Co., 1953. 510 p. \$6.25.

At long last we have a textbook in statistics for students of education which is non-mathematical, comprehensive, accurate, and up-to-date. Though the title of the Walker and Lev book indicates that it may be of interest to researchers in many different fields, the senior author is well-known as a professor of education and virtually all the material in it is written in the context of problems in educational research.

Although it may not be considered an elementary text, since by conventional standards it includes 'advanced topics' in statistics, it features the intuitive approach to the development of statistical concepts, an approach designed for the student not prepared to read mathematical literature. The first nine chapters are developmental and designed to cover the basic concepts of statistical inference. They include essential ideas of probability, population and sample, sampling distributions, chi-square, inferences concerning the mean and variance, and analysis of variance. The remaining chapters depend upon the theory introduced in the first part of the book in applications of statistical theory to such problems as correlation and regression, measurement and reliability, analysis of variance and co-variance. Included is an excellent final chapter on the very practical subject of non-parametric methods. An example of much needed modern content not hitherto available in books for students in education is a treatment of sample survey design.

In the first three chapters, through the medium of simple numerical problems, Walker and Lev 'intuitively' introduce the student to most of the statistical concepts he needs to understand statistical inference. This includes an introduction to probability and probability distribution (including standard symbolism for expressing probability statements), the notion of parameter versus statistic, the binomial distribution, probability density, mathematical expectation, the normal probability distribution, randomness in sampling, statistical estimation, the power function, and sampling from a finite population. Later chapters are unusually thorough, accurate, and complete on the subjects they cover.

The usefulness of the publication is enhanced by generally good illustrations and examples throughout, and an exceptionally comprehensive set of statistical tables and charts in the appendix. Good references are included for students who might be interested in further pursuit of topics.

Legitimate comments on this book must be so overwhelmingly in its favor that criticisms of it seem trivial. The writer of this review found the gains in the adoption of this text in an advanced course in educational statistics to outweigh by far the difficulties encountered with it. For a first edition there are a negligible number of typographical and numerical errors not covered in the errata sheet supplied by the publisher. For instance, a bar is left off of an X in equation (9.25), and what should be 77 appears as 73 in an equation at the bottom of page 214. These are of minor consequence in the use of this text. The writer found but one example which seemed to distract students from the statistical objectives. This appeared in Chapter 9 in an otherwise excellent development of the model for analysis of variance. It dealt with archery scores, the conceptualization of which seemed about as difficult for students to grasp as the concepts they were used to illustrate. This is an exception, however, since in general examples are realistic and well chosen.

The unique order of presentation of materials covered in this book may present some difficulties. This will depend upon how it is used. The introductory chapters emphasize models, theory, and concepts, and progressively and adequately build up a working background. However, the writer finds that inexperienced students find this somewhat tedious. Even in advanced courses education students appear to need an occasional opportunity to examine the application of a principle in a number of different ways. In one sense, it is logical to initiate such things as the power concept early in a course in statistical inference. From the standpoint of efficiency of instruction, however, this may profitably be delayed until after experience has been offered in some of the operations of statistical inference which does not require it. Related concepts of efficiency and consistency of estimates appear early in the text but are not clinched except in a supplementary chapter dealing with percentiles (or 'inefficient' statistics). Such matters, of course, are in the category of choices which usually are expected to be faced by an instructor as he

adapts to his class and his own objectives. The writer found it necessary, for instance, to supplement and emphasize equation (9.9) on the partitioning of sums of squares in a simple analysis of variance problem, an important basic concept appearing to need proof and elaboration. It is only fair to say, however, that there are undoubtedly many alternatives in arranging the sequence of a course in statistics and that it is too much to expect a text on this level to be self-teaching.

The authors in their preface suggest that some students might study the text without previous courses in statistics but that those without facility and quantitative thinking might profit by a "slow-paced introductory course". In most teachers colleges and colleges of education this text will probably be most useful in a second and/or a third semester course. As an advanced text it should be widely accepted. It will be found to be an efficient vehicle to the understanding of modern statistical methods in educational research. The material is there and the exposition is precise and clear.

FRANCIS G. CORNELL

University of Illinois

WILLIAM STEPHENSON. *The Study of Behavior. Q-Technique and its Methodology*. Chicago: University of Chicago Press, 1954, pp. 376. \$7.50.

This book does not, as the title implies, deal with behaviour, at least, not in the sense that the term is usually understood by psychologists; it deals essentially with techniques of obtaining self-ratings and ratings of others, called Q-sorts by Stephenson, and with techniques of correlating persons and factor analyzing the resulting tables of correlations, called Q-techniques by Stephenson. It constitutes, therefore, a summary and detailed statement of work carried on intermittently by Stephenson since 1935, and well known in its main aspects to statistically-minded psychologists. Those who have followed this work, and the many controversies to which it has given rise, will turn to this book in anticipation of finding a detailed discussion of the main points, together with an empirical proof of the effectiveness of the methods suggested.

Of the two innovations which Stephenson claims to have introduced, it may be said that the Q-sort presents a somewhat novel

and, under certain circumstances, a probably useful technique of self-rating. The subject is given a large number of statements similar to those appearing in orthodox questionnaires; he is then asked to rank these in order of applicability to himself, putting the one that characterizes him best at the top, and the one least like him at the bottom. The relationship between a Q-sort and an ordinary 'Yes'-'No' type of questionnaire appears to be essentially the same as that between a product-moment correlation and a tetrachoric correlation; though both do pretty much the same sort of thing, under certain circumstances the one may be preferable to the other. Stephenson never discusses in detail why he considers a Q-sort to be preferable to an ordinary questionnaire, and this reviewer, at least, found it difficult to see why claims made for the Q-sort should not be applied in equal measure to the despised questionnaire. On the evidence supplied by Stephenson, then, we can only conclude that the Q-sort, while ingenious, is unlikely to support the elaborate super-structure erected on it by Stephenson.

With regard to the technique of correlating persons, we find ourselves in the midst of a very stormy controversy. Most orthodox psychologists have pointed out that the techniques used by Stephenson are not new, that many of the detailed applications are subject to damaging statistical criticism, and that the results in cases where the method can legitimately be applied are no different from those which would have been reached by the more orthodox technique of correlating tests or questionnaire items. A good deal of Stephenson's book is taken up with attempts to answer these criticisms, and the reader is liable to get lost in these somewhat heated exchanges. By and large, this reviewer has come to the conclusion, after examining the evidence carefully, that the critics are right and that anyone wishing to use this technique would be well advised to discuss the methodological principles involved with someone expert in this field.

Stephenson does not help his case by showing a somewhat disingenuous tendency to change the meaning of the term Q-technique over the years, whilst pretending that what he now means by it is what he has meant by it all along. Simple examination of his earlier papers and the examples quoted therein is sufficient to show that his views have changed considerably. There

is no harm, of course, in changing one's views, but it does make controversy exceedingly difficult to follow when, quite justifiable, criticisms of earlier work are rebutted by referring to later and quite different definitions and interpretations.

Stephenson has, from the beginning, illustrated his discussions by means of small-scale experiments, not, presumably, meant to be taken very seriously and serving only illustrative purposes. In the present book, one might have hoped for a somewhat longer, detailed, and worked-out example of his technique, demonstrating precisely how it could be useful in advancing psychological theory and knowledge. Stephenson has not taken this opportunity, however, but has relied again on a large number of very small and often exceedingly trivial examples. It is difficult to judge a method in the absence of even a single instance where it has been applied to its best advantage. However, on its present showing it seems difficult to take very seriously the somewhat extravagant claims which Stephenson makes for it. The reviewer is willing to reverse his judgment if a convincing example could be given to demonstrate the usefulness of Q-technique. In view of the failure over a period of twenty years to provide such an example, however, it seems unlikely that this judgment will be reversed.

One last feature of the book calls for comment; namely, its style. A great deal of the argument is difficult to follow because of Stephenson's somewhat explosive and disjointed manner of writing. There is little in the way of sustained argument, but a constant switching of discussion from statistical detail to philosophical speculation, to attack on critics, to illustrative example, and back again to statistical detail. Distrusting his own reaction to this, the reviewer handed the book to several statisticians who confirmed his opinion that, from the statistical point of view at least, it was impossible to find any concise, consecutive, and properly documented demonstration of Stephenson's main claims.

In summary, there is much that is ingenious, interesting, and suggestive in this book, but it cannot be recommended to anyone not thoroughly familiar with the field and able to discount the more unorthodox and less acceptable parts of the argument.

H. J. EYSENCK

Maudsley Hospital, London

Drug Addiction among Adolescents. Sponsored by the Committee on Public Health Relations of the New York Academy of Medicine with the Assistance of The Josiah Macy, Jr. Foundation. New York. The Blakiston Co., 1953, pp. 320.

There was widespread alarm concerning the growth in the use of drug addiction in adolescents in the early fifties. In New York a number of committees were formed to better combat this evil. One of them was sponsored by the Welfare Council of New York City. A committee of fifty-eight decided to call a conference for the purpose of obtaining the thinking of authorities in medical, psychiatric, sociologic, educational, and law-enforcement fields. This was done with the help of the Josiah Macy, Jr. Foundation. This volume reports two sets of such conferences: a one-day conference held at the New York Academy of Medicine on November 30, 1951, reported in the first section of the book. The subcommittee held the second conference which lasted for two days without any prepared papers except a listing of subjects for discussion to be presented to the group. This was held at the New York Academy of Medicine on March 13 and 14, 1952, and the supposedly spontaneous discussion was recorded and is presented in the second section of the book.

One of the conclusions that came out of the first conference was a more alive awareness on the part of the participants that drug addiction among adolescents is primarily a sociological, psychological, and economic problem and only, secondarily, a medical one. There were some participants who emphasized the abnormal psychology or medical slant. But, generally, a realization that only a multiprofessional approach was at all promising seemed to hold sway. Almost every phase of the problem was considered; not in any thorough going fashion, however. Some of the problems concerned with were the need for elimination of concepts of moral depravity that doctors still have; the comparisons based on reminiscences of some of the participants with factors which were realized thirty years ago; the reason for the increase in adolescents' use of drugs; how the craving could be prevented; how the supply can be stopped; how to rehabilitate those who have been addicted; the rôle of education; what can be done in an institution; and what can be done in other directions including the education of parents. All of these and other related problems are considered these

conferences. One of the direct results of these two conferences was the inauguration of a project to correlate and integrate data on adolescent drug addiction gathered by the magistrates' courts of the city of New York. With the assistance of the United States Public Health Service, the project is now being carried on in New York University.

A small agenda committee might have prepared the material so that more alive and present-day efforts would be considered from the point of view of action for research and practice. But the Macy Foundation deserves congratulation on arranging for this conference and in helping in the publication of the materials of it. People in the field will not find this hard reading; they will find it segmented and piece-meal, however. Short comprehensive summaries would have helped clarify the presentation for the reader. The nearest thing to a feature of this kind shows up in the early portion of the second conference where Dr. E. H. L. Corwin is invited to summarize the first conference. For orientation purposes the conferences evidently served. The appendix contains a glossary of terms used from drug addiction—argot, this is called—and also an index. In a short but enlightened introduction by Dr. Frank Fremont-Smith, Medical Director of the Josiah Macy, Jr. Foundation, the purpose and significance of these conferences is expressed more coherently than in the other contributions. "What has developed," says he, "is a much clearer recognition that drug addiction in adolescents must be seen, not as a problem of moral degeneracy, not merely as a failure of law enforcement, but rather as one symptom of the serious deprivations suffered by many children living in large and crowded cities."

H. MELTZER

Psychological Service Center
St. Louis, Missouri

M. E. BENNETT. *College and Life*. New York: McGraw-Hill Book Co., 1952, pp. 457. \$4.50.

The author says that *College and Life* "deals with the various problems of learning and living in college, which each student must face and solve for himself as a democratic citizen in his college community in order to achieve the real values of a college education." Again, "Vocational guidance and mental hygiene form the warp and woof of the materials throughout the book . . ."

There are three parts to the book. In the first of these, entitled "Living in College," the general college scene is discussed from various points of view. The student is asked to consider seriously whether or not he is one of those who really should be in college. What the student may expect to get out of college in return for the time and money invested is discussed. Such aspects of college living as extra-curricular activities, customs and traditions of one's college, time budgeting, kinds of majors, trends in college curricula, importance of scholarship, participation in school politics as an effective college citizen, and learning self-direction come under consideration. The aim is stressed of achieving effective local, national, and international citizenship. Bennett says "College experience should provide every student with some excellent training in how to choose a leader, how to follow a leader, and how to be a leader." These are sound aims, too often neglected in the thinking of persons responsible for the training of students of all ages, and woefully absent in the thinking of the students themselves.

In Part Two, "Learning in College," four chapters are given to the nature and conditions of learning and retention, and to effective study habits. Chapter 8, the discussion of conditions for effective learning, would need implementation by way of reading or class discussion. The usefulness of the study-conditions and health inventories provided in Chapter 9 is limited by the student's probable lack of adequate criteria for answering. The exercises for improving memory in Chapter 10 may also lose in effectiveness from the student's inability to evaluate his faults and to outline changes in his study practices. The remaining five chapters of this section touch upon use of the library, effective silent reading, note-taking, thinking with a purpose, and the appraisal of progress in college. Generally, these chapters contain much of value to the beginning student. However, such exercises as the evaluation of reference materials call for knowledge and experience not possessed by many freshmen. The usual freshman would also need a great deal of help with the reading inventory, inasmuch as he simply would be unable to make some of the judgments required. The chapter on purposeful thinking seems quite good, and includes exercises which should be of value. The chapter on appraisal of progress should help to reduce the student's fear of examinations through understanding of their uses, and give him techniques for doing well in them, other things being equal.

Part Three, "Building a Life," aims at the development of personality, achieving mental health, developing a life philosophy, and achieving satisfactory adjustments in family life and in democratic citizenship. As is the case with the other sections, there is a lot here of value for the beginning student, but much of it is presented in rather a shallow manner, and the exercises for the student call for more experience and knowledge than he is likely to have. It would seem that the demands made upon the instructor, if the material is to benefit the student very much, are great.

It appears to this reviewer that throughout the book the exercises and indicated discussions would call for a counselor considerably better-trained than the usual 'mere' instructor. The somewhat elementary level of presentation of much of the material may be justified in view of its intended use with beginning college students, but adequate use of the material to bring about good adjustment and recognition and acceptance of the stated goals on the part of the students would require the services of a teacher well-trained in clinical psychology or counseling. If such a person were available, this book should provide a basis for an excellent orientation course. The 405-item bibliography contains excellent and recent references for students at all ability levels.

CECIL M. FREEBURNE

Bowling Green (Ohio) State University

LEOPOLD BELLAK, Ed. *Psychology of Physical Illness. Psychiatry Applied to Medicine, Surgery, and the Specialities*. New York: Grune and Stratton, 1952, pp. 243.

This is a very little book on a very big subject. Perhaps the magnitude of the field justifies the two titles carried by this work, but most certainly its contents do not. The subtitle is accurate, but the primary title is highly misleading. Of psychology, as psychologists know it, this book contains nothing. The last statement must be tempered only by recognition that the one chapter which appears to show some semblance of effort at scholarly review of existing literature (Chapter 5, on Psychiatric Problems in General Surgery) has one reference to a paper co-authored by a psychologist and dealing with a scientific, psychological study of determinants of post-surgery recovery. This paper receives no treatment in the text. It is simply given super-script reference together with three

other papers which, like the great bulk of the references used by the various authors, are clinical accounts of and superficial hypotheses about psychiatric phenomena associated with illness and physical defect.

This is a book by M.D's for M.D.'s. More specifically, it is a collection of brief papers by psychoanalytically trained or oriented medical specialists, physicians who obviously consider analytic concepts as both valid and efficient as a basis for 1) investigating, 2) understanding, and 3) controlling the doctor-patient relationship. In these papers, emphasis is upon understanding of the medical patient's psychological problems primarily within the framework of the abstractly descriptive terminology of psychoanalysis. Scientific investigation (as contrasted with clinical accounting) and recommendations as to techniques of management are generally given short shrift in preference to relatively stereotyped accounts of the 'meaning to the patient' of his particular symptom or defect. Much use is made of the 'body image' concept with little effort to outline its developmental psychological basis.

Of the possible and probable purposes which a book carrying the titles of this one might have, it is difficult to find that any one has been pursued. It is very obviously not a comprehensive review of psychological data or information concerning physical illness or handicap in the sense of Pintner, et al., *The Psychology of the Physically Handicapped*, or Barker, et al., *Adjustment to Physical Handicap and Illness*. As a matter of fact, the goal and effort represented in outcome in these two works is considered by the editor of this book as 'nearly impossible'. This book is obviously not a report of medical and psychological research carried on by a team of investigators concerned with any particular sample of patients, as exemplified by Keys, et al., *Biology of Human Starvation*. It is obviously not an attempt to review the different theories and systematic formulations of outstanding psychological and medical experts concerning psychiatric implications of illness.

What of the stated purpose of the book? According to the editor, it is offered "with the hope that we may contribute to a better medical practice—one which is more enjoyable for the doctor and more beneficial to the patient." The exact way in which this contribution is to be achieved is not stated explicitly. Certainly it is clear from the content that the contribution is not to be in the form of factual information (e.g., frequency of significant psychiatric complications

in surgical procedures). Nor is the contribution to be in terms of reasonably detailed recommendations concerning the management of the emotionally involved patient (e.g., possible techniques for effecting a psychiatric referral of a patient with multiple sclerosis). Nor is this book's contribution to a 'better medical practice' to be in the form of challenging questions and the mapping of unexplored areas. This book raises no questions and in no way suggests that there are any areas of 'psychology' in medical practice that require objective investigation. So far as may be determined from this collection of little essays, the 'psychology of physical illness' is that contradiction in terms—a finished science.

What of the actual content? This is an edited collection of fifteen papers by sixteen authors. The medical areas covered include general practice, internal medicine, malignancy, general surgery, plastic surgery, gynecology and obstetrics, urology, orthopedics, eye, ear, nose and throat, neurology, pediatrics, dermatology, dentistry. The final chapter is on the "Personality of the Physician as a Factor." In view of the pervasive psychoanalytic orientation of the book, one can well imagine the editor's frustration at having to dispense with a chapter on proctology because he was not acquainted with any expert in the rectal field who was "at the same time competent in psychiatry."

With the exception of the editor, the writers are not nationally prominent in either psychology or psychiatry. This may account for the brief biographical sketch which precedes each chapter and helps to establish for the reader the special competence of the particular writer. The biographical similarity of the writers is most striking. All sixteen are residents of New York City. With the exception of two private practitioners, the remaining fourteen contributors have formal affiliation with one or another of five prominent New York medical institutions. Half of them are associated with either Mount Sinai Hospital or Columbia University. Seven of the authors are formally affiliated with the New York Psychoanalytic Society. Only half of the authors record any previous publications in the references appended to their chapters. The chapter reference lists which range from six to fifty-eight items in length, and average nineteen items, are somewhat expansively termed 'bibliographies'.

The general impression left with this reader by both the chapter

contents and the biographical data on the writers is that their primary qualifications consisted of convenient geographical accessibility to the editor and a happy homogeneity of philosophical (i.e., psychoanalytic) conviction. The collaboration and editorial direction have resulted in a work of surprisingly even quality in contrast to the very marked variation of style, organization, emphasis, and thoroughness which is a common criticism of similar compendiums. Unfortunately, the desirable evenness of quality is achieved by a failure on the part of the participants to raise themselves above the platitudinous level of observations "such as" a person has 'feelings' about his sickness and that it is important for the physician to recognize and deal with these feelings. The chapter on dentistry, for example, is lush with unqualified statements about how the "usual patient feels about his mouth," about tooth decay, and about dental therapeutics, but there is no indication of even an awareness on the part of the author of the extensive research literature on psychological determinants of individual differences in pain reaction. 'Pain' has but a single reference in the index of this book and the term 'limen' appears not at all. The one chapter which stands above all the others because of its unhackneyed observations, the subtlety of its insights into the implications of established facts of psychology, and its emphasis upon the importance of the patient's information and control of a medical procedure as an anxiety alleviant, is the chapter on general surgery. Typical of this writer is his recommendation that operating room personnel should "be aware of the fact that during the induction of general anesthesia, hearing is the last sense category to be obliterated and that profound impressions can come by way of the partly narcotized hearing apparatus."

Books whose authors and publishers assign titles which necessarily suggest appropriate comparison with other similarly titled works, or whose titles arouse natural expectations as to content, must bear the test of such comparisons and expectations. A comparison of this little book with a big title against such works as those of Pintner and Barker indicates differences both of goal and achievement of the magnitude of the differences between a Sunday picnic and the French expedition to the top of Annapurna. In view of the strongly critical flavor of this review and the current revival of 'book-burning' one might ask if this is a dangerous or harmful

book. It is not. It may be perused safely by medical students, general practitioners, and specialists—even by the physically-ill person. It is unlikely to add to the knowledge or understanding of any of these, but it most definitely will not harm them. It is not that good.

WILLIAM SCHOFIELD

● *University of Minnesota*

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 45

November, 1954

Number 7

A FACTOR ANALYSIS OF STUDENT RATINGS OF PSYCHOLOGY INSTRUCTORS ON THE PURDUE SCALE

A. W. BENDIG

University of Pittsburgh

In a previous study of the use of student rating scales as diagnostic and evaluative instruments to assess instructional effectiveness at the college level it was suggested that ratings from scales that are highly interrelated might be combined to yield a fewer number of indices of teaching proficiency that would be more reliable than single scale ratings (1). Coffman (4) presents a similar point of view, arguing that a multiplicity of single scales presents difficulties in interpretation. In addition, results have indicated that a simple summing of ratings from different scales, without consideration to varying intercorrelations among pairs of scales, can obscure important information. The residual variance remaining in achievement grades in a course in introductory psychology after the variability attributable to academic aptitude had been eliminated was shown to be significantly related to certain single scales on the Miami Instructor Rating Sheet and to be unrelated to others. A simple combining of the ratings resulted in a loss in the relationship between student ratings and residual achievement (7).

Several methods could be used to combine ratings from different scales. One would be to combine scales that rationally appear to measure different facets of a single broad characteristic. The results mentioned above (7) suggested a combining of scales measuring only characteristics of the course and a similar combining of scales related to characteristics of the instructor. A simple summing of these two types of scales was used in a study of the rela-

TABLE 2.—ORIGINAL AND ROTATED FACTORS FOUND ON THE PURDUE RATING SCALE FOR INSTRUCTION

P.R.S.I. Scale	Original Factors			Rotated Factors			h ²
	I	II	III	I	II	III	
1	65	39	33	29	77	07	68
2	78	-48	-28	66	-02	70	93
3	54	-64	07	22	00	81	70
4	58	14	-58	83	00	00	69
5	87	46	23	52	86	10	102
6	90	24	-32	89	40	1	88
7	58	48	31	27	78	-05	68
8	72	-47	-09	49	09	71	75
9	57	-53	36	06	27	81	73
10	77	51	07	58	72	-05	86
Per cent of variance	50	21	9	30	27	23	80

only 0.02. Extraction of factors was terminated after three factors had been extracted, since an inspection of the residuals plus the application of the criteria suggested by Vernon (9, p. 130) and Tucker (3, p. 298) indicated that the residual variance lacked significance.

The obtained factor loadings were graphically and analytically rotated to achieve orthogonal simple structure. The original and rotated factor loadings can be found in Table 2. Rotated Factor I appears to be a general factor with medium or high loadings on all P.R.S.I. scales with the exception of Scale 9. This appears to be a broad 'halo' factor that is common to most of the P.R.S.I. scales.

Rotated factors II and III showed a very clear simple structure. Factor II had high loadings on the following four P.R.S.I. scales.

Scale	Description	Loading
5	Presentation of subject matter	0.86
7	Self-reliance and confidence	0.78
1	Interest in subject	0.77
10	Stimulating intellectual curiosity	0.72

The above four scales seem to hang together on rational grounds. All of them refer to aspects of the instructors' behavior involved

in his performance as a lecturer and only as a lecturer. In naming this factor (always a subjectively hazardous task), we might suggest the appellation of 'Instructional Competence'.

Four other P.R.S.I. scales showed high correlations with Factor III:

Scale	Description	Loading
3	Fairness in grading	0.81
9	Personal appearance	0.81
	Personal peculiarities	0.71
2	Sympathetic attitude toward students	0.70

Again a common thread runs through the scales loaded with this factor. The theme appears to be the personality of the instructor as viewed by the student. The personal interaction of instructor and student would seem to be important here: what impression that the student has of his instructor's objectivity, empathy, and personal characteristics (both structurally and functionally). If forced to, we should prefer to call this factor 'Instructor Empathy'.

One comparable study is the factor analysis of the P.R.S.I. reported by Smalzried and Remmers (8). Since this previous study analyzed the data from a high-school sample and only extracted

TABLE 3.—COMPARISON OF FACTOR LOADINGS ABOVE 0.50 FOUND IN TWO STUDIES OF THE PURDUE RATING SCALE FOR INSTRUCTION (DECIMAL POINTS OMITTED IN TABLE)

P.R.S.I. Scale	Smalzried and Remmers		Present Study	
	Factor I	Factor II	Factor II	Factor III
1	76		77	
2		77		70
3		88		81
4		73		
5	89		86	
6	69			
7	89		78	
8	72	53		71
9		79		81
10	67		72	

two factors from the matrix of intercorrelations (militating against a simple structure solution), it could not be expected that the results of Smalzried and Remmers would be comparable to the present results. However, an inspection of their data suggested an interesting comparison. In Table 3 are given the high loadings (above an arbitrary figure of 0.50) for the two factors found by Smalzried and Remmers and also the high loadings for rotated factors II and III in the present study. A remarkable parallelism is shown in the two sets of data. Smalzried and Remmers found more high loadings for their factors I and II than did the present study for factors II and III. A clue is offered to explain the discrepancy from the data in Table 2. P.R.S.I. scales 4 and 6, for which Smalzried and Remmers found high loadings and the present study did not, are the two scales with the highest loadings on our general factor I. Since this general factor was not separately extracted by Smalzried and Remmers, its variance is confounded with the variances of the other two important factors and partially obfuscates their structure. It is suggested that the extraction of a third factor by Smalzried and Remmers would have given, after subsequent rotation to simple structure, a factor structure for the P.R.S.I. remarkably similar to the one found here.

The scale factor loadings in Table 2 can be looked upon as the correlation of each P.R.S.I. scale with a 'pure' measure of a particular factor. In a sense they are validity coefficients useful in predicting the criterion of a factor score for an instructor. From these loadings, and the intercorrelations of the scales reported in Table 1, multiple regression equations for predicting factor scores were derived by the standard method outlined by Peters and Van Voorhis (6, pp. 225-229). Scales 1, 5, and 7 were found to have a multiple correlation of 0.89 with factor II and a similar correlation was found for P.R.S.I. scales 3, 8, and 9 in predicting factor III scores. Both coefficients were reduced to 0.86 by the shrinkage formula given by McNemar (5, p. 161). Table 4 gives the beta (b) weights for each of the P.R.S.I. scales, the raw score (B) regression weights, and the correction term for regression equations used in predicting factor II and factor III scores. In determining a factor II score for any instructor, his mean ratings on P.R.S.I. scales 1, 5, and 7 were multiplied by the B weights given in Table 4 and the correction added to the sum of these products. It must be

TABLE 4.—REGRESSION WEIGHTS USED IN PREDICTING FACTOR SCORES

P.R.S.I. Scale	Factor II		Factor III	
	b	B	b	B
1	0.29	0.0377		
2				
3			0.31	0.0795
4				
5	0.38	0.0277		
6				
7	0.31	0.0360		
8			0.23	0.0397
9			0.47	0.0644
10				
Correction	-7.79		-16.10	
Multiple correlation:				
Obtained	0.89		0.89	
Corrected for shrinkage	0.86		0.86	

noted that what we hereafter shall call 'factor scores' are only estimates of the score an instructor would receive if we had a pure measure of the factor, but that these estimated scores should correlate to the extent of 0.86 with 'true' factor scores.

An interesting question raised was whether the objective tests that nine of the eleven instructors had taken upon entering graduate school could be used in predicting their factor scores. Three tests had been used in this entrance battery: (a) the Miller Analogies Test, (b) a test of mathematical ability, and (c) a reading comprehension test containing three subtests measuring vocabulary, reading speed, and level. The two factor scores estimated for each instructor were correlated with these test scores by both product-moment and rank-difference (ρ) methods. The correlation of each factor with the entrance tests, plus the intercorrelations of the factors, can be found in Table 5. The vocabulary subtest from the reading comprehension test correlated positively and significantly with both factors II and III, but this result needs investigation in another sample before such a test could possibly be used as a selection device for instructors of introductory psychology.

TABLE 5.—PRODUCT-MOMENT AND RANK-DIFFERENCE CORRELATION COEFFICIENTS BETWEEN RAW SCORES ON GRADUATE SCHOOL ENTRANCE TESTS AND ESTIMATED FACTOR SCORES FOR NINE INSTRUCTORS

Test Variables	Product-Moment		Rank-Difference	
	Factor II	Factor III	Factor II	Factor III
Miller Analogies	0.30	0.19	0.29	0.04
Mathematical Ability	-0.28	0.50	0.02	0.40
Reading Comprehension				
Vocabulary	0.77*	0.68*	0.81**	0.81**
Speed	0.55	0.33	0.51	0.48
Level	0.45	-0.42	0.37	-0.49
Factors (Intercorrelation)	0.19		0.32	

* Significant at the .05 level of confidence.

** Significant at the .01 level of confidence.

DISCUSSION

The major purpose of this study was to find whether factor analytic techniques applied to student ratings on the P.R.S.I. could provide an empirical method of combining separate scales to yield a smaller number of reliable and independent measures of instructor effectiveness. For this method to prove successful a clear and orthogonal simple structure solution must be found. The results reported in Tables 2 and 3 lead to the conclusion that these requirements are met by the factorial structure of the P.R.S.I. Eighty per cent of the variance of mean ratings on the first ten scales of the P.R.S.I. are accounted for by three orthogonal factors: a general factor permeating most of the scales and two group factors measuring independent dimensions of the teaching situation.

It must be acknowledged that only a relatively small number of instructors in a single subject matter area were used in the analysis, although the number of student raters participating was quite sizeable. However, the similarity in factor structure of the P.R.S.I. in this and in the only other published report suggest that the structure found here is neither a function of the smallness of the sample of instructors nor of using instructors from a single academic department. We should predict that

another sample would yield slightly different beta weights from those reported in Table 4, but the essential factor structure of the P.R.S.I. would remain unchanged.

The similarity of the factor structure of the P.R.S.I. found in this study to the structure previously reported by Smalzried and Remmers (8) has already been noted. In addition, Coffman (4) has reported a factor analysis of nineteen instructor ratings scales and the identification of four factors. Two of these four factors, Empathy and Organization, appear to be similar to the two group factors found in our studies. Coffman's third factor, Academic stereotype, may be similar to our general factor.

The practical usefulness of the results lies in the stimulus for future research. We feel that we now have two fairly pure, independent, and quantifiable measures of instructor effectiveness based upon student ratings whose relationships with other measures of teaching aptitude and ability can be investigated. A tentative, but suggestive, beginning is reported in Table 5. Future research will be concerned with (a) the relation of these factors to other criteria of teaching competence, (b) the effect upon these criteria of characteristics of the student (age, sex, aptitudes, achievement, etc.), and (c) the development of aptitude measures useful in predicting a multiple criterion of teaching competence in psychology.

SUMMARY

The mean ratings of eleven introductory psychology instructors on the first ten scales of the Purdue Rating Scale for Instruction were intercorrelated and the resulting matrix factor analyzed by a centroid method. Three factors accounted for eighty per cent of the variance: a general factor permeating the scales, and two group factors that were labelled 'Instructional Competence' and 'Instructor Empathy'. Regression equations were developed to estimate factor scores of instructors on these last two dimensions and in both cases three P.R.S.I. scales were found to show a multiple correlation of 0.86 with the factors. Nine instructors had taken a battery of graduate school entrance examinations and their test scores were correlated with their factor scores. One test, reading vocabulary, was found to correlate significantly with both factor scores.

REFERENCES

- 1) A. W. Bendig. "A comparison of psychology instructor and national norms on the Purdue Rating Scale for Instruction." *J. educ. Psychol.*, 1953, 44, pp. 436-441.
- 2) ———. "The effect of level of achievement upon students' instructor and course ratings in introductory psychology." *Educ. psychol. Measmt.*, 1953, 13, pp. 437-448.
- 3) R. B. Cattell. *Factor Analysis*. New York: Harper, 1952.
- 4) W. E. Coffman. "Determining students' concepts of effective teaching from their ratings of instructors." *Amer. Psychologist*, 1952, 7, 369.
- 5) Q. McNemar. *Psychological statistics*. New York: Wiley, 1949.
- 6) C. C. Peters and W. R. Van Voorhis. *Statistical Procedures and Mathematical Bases*. New York: McGraw-Hill, 1940.
- 7) H. E. Russell and A. W. Bendig. "An investigation of the relation of student ratings of psychology instructors to their course achievement when academic aptitude is controlled." *Educ. Psychol. Measmt.*, 1953, 13, pp. 626-635.
- 8) N. T. Smalzried and H. H. Remmers. "A factor analysis of the Purdue Rating Scale for Instruction." *J. educ. Psychol.*, 1943, 34, 363-367.
- 9) P. E. Vernon. *The Structure of Human Abilities*. New York: Wiley, 1950.

VALIDITY OF TEACHERS' RATINGS OF ADOLESCENTS' ADJUSTMENT AND ASPIRATIONS¹

DAVID P. AUSUBEL, HERBERT M. SCHIFF, AND

MAJORIE P. ZELENY

Bureau of Education Research, College of Education, University of Illinois

This study was concerned with the relationship between teachers' ratings of adjustment and aspirational traits in adolescents and various other self-report, objective, and projective measures of the same characteristics.

Teachers' ratings are frequently used for diagnostic, prognostic, guidance, and research purposes because of their ready accessibility and quantifiability. Moreover, when employed as averages of several raters and in relation to traits (a) which are overtly manifest and objectively observable in everyday behavior and (b) which can be specifically and unambiguously defined (8, 20, 21), such ratings enjoy considerable reliability. A further advantage of teachers' ratings is the fact that they can be based on frequent and prolonged observation of behavior under a variety of conditions.

With respect to teachers' ratings of adjustment, the main problem was to ascertain the kinds of relationships that prevail between these observational appraisals of personality traits involved in adaptation to the school environment and such other criteria of adjustment as absence of overt behavioral symptomatology or relatively low degree of deviancy with respect to the more general personality trends measured by projective instruments. To what extent do these different instruments measure similar, overlapping, or even entirely discrete aspects of the global concept of adjustment? A secondary problem had to do with the bases on which teachers' ratings of adjustment are made. To what extent are they related to ratings of aspirational traits and to pupils' scholastic competence and sociometric status?

¹ This research was supported by grants from the Bureau of Educational Research and from the University Research Board of the University of Illinois.

The second major problem dealt with in this study was the relationship between teachers' ratings of aspirational traits and 'real life' measures of level of aspiration. In a sense this question raises the problem of the validity of such ratings.

METHOD

Population.—The subjects in this study consisted of fifty students comprising, with the exception of a few absentees, the entire junior class of University High School² in Urbana, Illinois. The mean age of these students was 15.8 years, and the distribution by sex was twenty-four boys and twenty-six girls. A majority of students in this school come from professional homes, the parents generally holding academic appointments at the University of Illinois. Admission to the school, however, is unrestricted except for the payment of a small nominal tuition fee.

Teachers' ratings.—Five current teachers of each student were asked to make ratings at the end of the school year. In all, the ratings of fifteen teachers (eight male and seven female) were involved. Students were rated on a five-point scale on general adjustment (defined in terms of personality integration and emotional stability) and on three aspirational traits (persistence, scholastic competitiveness and academic aspiration). To minimize 'halo effect', the teachers were requested to rate all students on a single trait before proceeding to consideration of the next trait. The ratings of five teachers were averaged to provide a mean rating for each student on four traits.

'Real life' measures of aspiration.—'Real life' indices of aspirational level included both academic and vocational measures. The former consisted of (a) academic goal discrepancy—the algebraic difference between the student's expressed academic aspiration for the current school year and his composite grade point average for the preceding four semesters; and (b) academic performance estimate—the algebraic difference between the student's estimate of his cumulative high-school standing and his actual four-semester grade point average. Prediction of academic standing was made on a six-point scale (well below average to upper few), and aca-

² The authors acknowledge the coöperation and administrative assistance of Professor Charles M. Allen, Principal, and of the staff of University High School in collecting these data.

demic aspiration was expressed in terms of a five-point scale (get by to upper few). Grade point average was computed from actual grades which in this school are officially recorded on a five-point scale but are not divulged to students.

Three measures of level of vocational aspiration were utilized: (a) vocational prestige needs—the student's total weighted score on the level of interests, part of the Lee-Thorpe Occupational Interest Inventory³ (15); (b) vocational tenacity—a composite standard score on a hypothetical level of aspiration test consisting of three hypothetical situations in which the subject is successively requested to suppose that he is preparing for a career in medicine, engineering, and a skilled trade, respectively, and that in the course of his vocational preparation he meets serious obstacles of a specified nature. In each of these situations, four alternatives are presented ranging from maintenance of the original goal at all costs (high vocational tenacity) to complete abandonment of the original vocational goal (low vocational tenacity). The subject's choice in each situation, appropriately weighted, is converted into a standard score, and all three standard scores are added to yield a composite score. (c) vocational unreality—the total discrepancy between each of the subject's nine percentile scores on the Kuder Preference Record (14) and the corresponding mean percentile scores of persons in his occupation of choice as given in the manual for this test.

Measures of adjustment.—In addition to teachers' rating of adjustment, the following indices of adjustment were employed: (a) M.M.P.I. adjustment score—total adjustment score on the Minnesota Multiphasic Personality Inventory (11). In deriving total adjustment scores, raw scores on each of the subscales were first converted into standard scores. Since a positive standard score on a given subscale indicated a degree of deviancy with respect to the trait measured that was greater than the mean score characterizing our population, a measure of total personality deviancy could be computed for each subject by merely summing

³ This is a measure of general level of job prestige desired by an individual. The subject is presented with thirty triads of occupational activities in each of which area of work involved is held relatively constant while level of job prestige is varied. Total score is computed from the subject's preferences in each triad appropriately weighted in terms of job prestige.

all of his positive standard scores on the various subscales.⁴ Finally, by reversing the sign of the total standard score, a measure of adjustment was obtained in which a high score was indicative of relatively good adjustment. (b) Rorschach adjustment score—an adjustment score derived from the Rorschach group test administered in accordance with the procedure developed by Harrower and Steiner (10), and scored in accordance with the system developed by Klopfer and Kelley (13). The adjustment score represents the total number of entries made in Munroe's check list⁵ (18) expressed as a standard score with the sign reversed so that a high score is indicative of good adjustment. (c) Rorschach anxiety score—the total number of occasions specified signs of anxiety appeared in the subject's protocol. Criteria of anxiety included fifteen commonly accepted signs such as diffuse shading responses, signs of shading shock, high number of card rejections, threatening or evasive responses, high percentage of oligophrenic responses, etc. (9, 13, 19). Each time one of these fifteen anxiety signs occurred, the subject was credited with one point. Thus, high Rorschach anxiety scores are indicative of relatively high levels of anxiety and vice versa. (d) Composite adjustment score—the algebraic sum of the subject's standard M.M.P.I., Rorschach adjustment and teachers' adjustment rating scores. (e) I.P.I. anxiety score—total score on a paper-and-pencil inventory designed to measure anxiety level (Illinois Personality Inventory). (f) Sociometric status—the mean sociometric rating earned by each student when rated sociometrically by all of his classmates on a five-point scale ('definitely not wanted as a friend' to 'wanted as a best friend'). (g) Grade point average—the student's cumulative grade point average over four semesters computed on a five-point scale. As in the case of sociometric status, a numerically high score is at the desirable end of the scale.

⁴ Since a negative standard score is indicative of deviancy on the masculinity-femininity subscale, negative standard scores on this scale were added to the other standard scores to obtain total M.M.P.I. adjustment score (prior to final reversal of sign).

⁵ This check list proposes criteria for a 'normal' range of responses for each type of scoring determinant. Entries in the check list, therefore, represent deviancy from the proposed limits of normalcy.

TABLE 1.—INTERCORRELATIONS AMONG PERSONALITY TRAITS AS RATED BY TEACHERS

	Adjustment	Persistence	Scholastic Competitiveness	Academic Aspiration
Adjustment	.90*	.71	.60	.42
Persistence	—	.87*	.81	.73
Scholastic Competitiveness	—	—	.83*	.79
Academic Aspiration	—	—	—	.90*

NOTE: All coefficients of correlation in this table are significant beyond the one per cent level of confidence.

* Mean intercorrelation between five teachers' ratings of subjects on same trait computed by 'squared r ' method, and corrected by Spearman-Brown formula to indicate the predicted reliability of the combined judgments of five raters.

RESULTS AND INTERPRETATION

Reliability and Generality of Measures.—The reliability (generality over-raters) of teachers' ratings is shown in Table 1. For each trait the mean intercorrelation (computed by the squared r method) between the ratings of five teachers on the same pupils was obtained. This was corrected by the Spearman-Brown formula to indicate the predicted reliability of the average rating of five teachers. Application of the correction formula was considered justifiable since an average of five ratings is more reliable than a single rating in the same sense and to the same predictable degree that a ten-item test is more reliable than a two-item test. The resulting reliability coefficients approximate the range of magnitude generally obtained with combined teachers' ratings.

Table 1 also shows that teachers' ratings exhibit considerable generality over traits. This is not necessarily indicative of 'halo effect' since the three aspirational traits both closely resemble each other and are not logically unrelated to adjustment. Furthermore, the intercorrelations among aspirational traits are higher than the correlations between these latter traits and adjustment.

The reliability⁶ of academic goal discrepancy and of academic

⁶ A more definitive presentation and discussion of the reliability, generality, and psychological significance of 'real life' measures of level of aspiration are given in a previous paper (5). The more relevant findings are summarized here to provide some objective basis for appraisal of these measures.

performance estimate scores could not be ascertained, but there was a high degree of relationship between these two different indices of level of academic aspiration. Vocational tenacity scores on the three different kinds of situations (i.e., medicine, engineering, and skilled trade) exhibited a degree of generality which was proportionate to the degree of similarity between the vocations. The test-retest reliability of the level of interests scores on the Occupational Interest Inventory (identical with our measure of vocational prestige needs) is given as .88 in the test manual (15). The reliability of vocational unreality scores could not be determined. Vocational prestige needs were significantly related to both vocational unreality and vocational tenacity in boys, but correlations between academic and vocational measures of level of aspiration were not significant.

It was not feasible for either technical or administrative reasons to obtain reliability coefficients for the Rorschach measures, the M.M.P.I. adjustment score, or the composite adjustment score. Intercorrelations among Rorschach adjustment score, M.M.P.I. adjustment score and teachers' rating of adjustment (see Table 2) were generally non-significant except for the correlation between the latter two measures in our female subjects. The correlations between each of these three measures and composite adjustment score were spuriously high because each measure was a component part of the composite score. M.M.P.I. adjustment score was the most representative component of the composite measure in boys, and teachers' rating of adjustment enjoyed comparable status among girls.

TABLE 2.—INTERCORRELATIONS AMONG ADJUSTMENT SCORES

	Rorschach Anxiety		M.M.P.I. Adjustment		Teachers Adjustment		Anxiety Level (I.P.I.)		Sociometric Status		Composite Adjustment	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Rorschach Adjustment	-.24	-.19	.31	-.21	-.26	-.09	*-.40	-.32	-.17	.30	**59	.40
Rorschach Anxiety	—	—	-.11	.01	-.21	-.16	.36	.06	-.17	-.22	-.21	-.20
M.M.P.I. Adjustment	—	—	—	—	.05	**57	-.06	**56	.04	.13	**79	**61
Teacher Adjustment	—	—	—	—	—	—	.03	**52	**75	**64	.40	**79
Anxiety Level (I.P.I.)	—	—	—	—	—	—	—	—	.03	.01	.17	*.45
Sociometric Status	—	—	—	—	—	—	—	—	—	—	.44	**59

* Significant at the five per cent level.

** Significant at the one per cent level

Thus, it appears that the aspects of adjustment measured, respectively, by teachers' observational ratings of classroom behavior, a self-report index of behavioral deviancy with respect to various syndromes in psychopathology, and a projective measure of general personality integration are relatively unrelated to each other in our sample of adolescents. If all three measures are equally valid, they must all be measuring quite different aspects of adjustment which (except for M.M.P.I. and teachers' ratings in girls) show no significant overlap whatsoever. The same type of relationship holds true for self-report (I.P.I.) and Rorschach measures of anxiety which were not significantly correlated. This latter datum is not in accord with the previously reported finding of a significantly positive relationship between these two measures of anxiety in a population of college students (4).

More in accordance with expectations, I.P.I. anxiety scores correlated negatively with composite and M.M.P.I. adjustment scores, and with teachers' ratings of adjustment in girls, and with Rorschach adjustment scores in boys. The relationships between Rorschach anxiety scores and other measures of adjustment were not significant. At least among our adolescent girls, therefore high levels of anxiety as measured by the I.P.I. tended to be associated with poor adjustment and vice versa. A possible explanation for this sex difference might lie in the prevailing cultural attitude that high levels of striving (and of the anxiety that accompanies such striving) are more acceptable and compatible with good adjustment in boys than in girls. In support of this interpretation was the fact that teachers' adjustment ratings were more consistently negatively related to boys' than to girls' academic goal discrepancy scores and academic performance estimates (see Table 4).

Relationship between teachers' ratings and other measures of adjustment.—As already noted, teachers' rating of adjustment was not significantly related to Rorschach adjustment score in either sex, but was positively related to M.M.P.I. adjustment score in girls ($p. < .01$). Teachers' rating of adjustment was also significantly more highly related to composite adjustment score in girls than in boys, although in each case the coefficient of correlation was spuriously high. Rorschach anxiety score was not significantly correlated with teachers' rating of adjustment, but

I.P.I. anxiety score was negatively related to the latter score in girls ($p. < .01$).

It would appear, therefore, that for our population of adolescents, teachers' ratings of adjustment are a more valid and psychologically meaningful measure in relation to girls than in relation to boys. Appraisal of boys' adjustment status is very possibly complicated in the minds of teachers by the fact that certain traits related to extremely high levels of striving (and the associated anxiety) are regarded as unfavorable for adjustment by mental hygienists, but are nevertheless accepted as desirable in connection with our cultural ideal of masculine success. In the case of girls, on the other hand, teachers are apparently less confused by contradictory criteria of adjustment. Supporting this interpretation are (a) the significantly negative correlation between girls' I.P.I. anxiety scores and teachers' adjustment ratings, in contrast to the approximately zero coefficient of correlation between the corresponding scores of boys, and (b) the more consistent tendency for teachers' ratings of adjustment to be correlated negatively with objective indices of academic level of aspiration in the case of girls (see Table 4).

A clue regarding the bases on which teachers' ratings of adjustment are made appears in the moderately high degree of positive relationship found between these ratings and (a) ratings on aspirational traits, (see Table 1), and (b) pupils' sociometric status and grade point average (see Table 3). The composite portrait of a pupil who receives a high teachers' rating on adjustment is an individual who is perceived by teachers as persistent and scholastically competitive, who is highly accepted by his classmates and who has a superior scholastic record. Table 4 also shows that his academic aspirations tend to be more commensurate with his past academic performance as evidenced by the negative correlation between academic goal discrepancy and teachers' adjustment rating. In addition, if the pupil is a girl, she will tend to have high M.M.P.I. and composite adjustment scores and a low I.P.I. anxiety score.

Table 3 also shows the relationships between teachers' ratings of aspirational traits and other non-rating scale indices of adjustment. Aspirational ratings were not significantly related to M.M.P.I. and Rorschach adjustment scores or to Rorschach

TABLE 3.—CORRELATIONS BETWEEN TEACHERS' RATINGS AND VARIOUS MEASURES OF ADJUSTMENT

Measures of Adjustment	Teachers Ratings							
	Adjustment		Persistence		Scholastic Competitiveness		Academic Aspiration	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
M.M.P.I. Adjustment	.05	** .57	-.06	-.02	-.01	-.05	-.16	-.24
Rorschach Adjustment	-.26	-.09	-.03	-.14	-.28	.22	-.12	.30
Rorschach Anxiety	-.21	-.16	-.19	.02	-.14	-.10	-.12	-.17
Anxiety Level (I.P.I.)	.03	** -.52	-.30	** -.52	-.38	-.20	-.30	-.34
Composite Adjustment	* .40	** .79	* .44	** .57	.23	** .53	-.04	** .70
Sociometric Status	** .75	** .64	** .54	* .48	* .46	* .42	* .40	* .39
Grade Point Average	** .59	** .61	** .78	** .72	** .92	** .84	** .61	** .71

* Significant at the five per cent level.

** Significant at the one per cent level.

anxiety ratings. A puzzling negative relationship ($p. < .01$) prevailed between girls' I.P.I. anxiety scores and teachers' ratings on persistence. Teachers' ratings of aspirational traits were significantly correlated with composite adjustment in girls; in the case of boys this was only true with respect to persistence.

In general, as one might expect, grade point average was more highly correlated with aspirational ratings than with ratings of adjustment, but the reverse held true for sociometric status (see Table 3).

Relationship between teachers' ratings and 'real life' measures of aspiration.—Table 4 shows that there is a pronounced tendency for academic goal discrepancy scores to be negatively correlated with such highly related teachers' ratings as scholastic competitiveness, persistence, and academic aspiration. Thus, students who expressed relatively high academic aspirations for the future in the light of previous scholastic performance tended to receive ratings from teachers which were indicative of relatively low aspirational level. This same tendency prevailed with respect to academic performance estimate, but was significant only for girls, and in only two out of three instances. Generally speaking, relationships between teachers' ratings of aspirational traits and

TABLE 4.—CORRELATIONS BETWEEN TEACHERS' RATINGS AND 'REAL LIFE' MEASURES OF LEVEL OF ASPIRATION

'Real Life' Measures of Level of Aspiration	Adjustment		Persistence		Scholastic Competitiveness		Academic Aspiration	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Academic Goal	*-.43	*-.49	**-.57	**-.59	**-.52	**-.57	-.30	*-.44
Discrepancy								
Academic Per- formance Es- timate	-.20	*-.46	-.20	*-.43	-.38	*-.48	-.15	-.31
Vocational Pres- tige Needs	-.25	-.10	.16	.11	.31	.23	.35	.25
Vocational Tena- city	-.11	-.19	-.04	-.15	.22	*.45	.34	.37
Vocational Un- reality	-.31	-.22	-.13	.37	-.08	.27	-.03	.33

* Significant at the five per cent level.

** Significant at the one per cent level.

measures of vocational aspiration were not statistically significant and exhibited no consistent trend with respect to direction.

This negative relationship between 'real life' measures of academic aspiration and teachers' ratings of aspirational traits (ostensibly based on motivational behavior in an academic setting) casts doubt upon the validity of the latter measures. It confirms findings in a previous study with intellectually gifted sixth-graders (1) in which the writer failed to obtain significant relationships between teachers' ratings of aspirational traits and a measure of prestige motivation (i.e., relative tendency to respond to an incentive of personal prestige by increasing work output over the level achieved under conditions of anonymity). Another supportive finding in this latter study was a zero correlation between these same ratings and pupils' self-ratings on aspirational traits (1). Lewis (16), on the other hand, obtained very significant differences between 'educationally retarded' and 'educationally accelerated' gifted children aged nine to fourteen with respect to teachers' ratings on perseverance and ambitiousness.

Although educational performance (holding other factors constant) is perhaps a more definitive criterion of the validity of ratings of aspirational traits than either self-ratings or objective

test measures, it should be realized that such evidence of validity is largely circular when the ratings are made by teachers. Since teachers' ratings of motivation are substantially based on academic performance, the latter criterion can hardly be considered independent. In the present study, correlations of .61 to .92 were obtained between teachers' ratings of aspirational traits and grade point average (see Table 3).

Taking all of the evidence into consideration, there is certainly good reason for questioning the validity of teachers' ratings of pupils' aspirational characteristics. This lack of validity is apt to be even more serious in the case of adolescent pupils, judging from the findings of the present study and from the relatively poor insight of teachers into high-school pupils' interests (6) and sociometric status (7). Also, at least with respect to the latter function, teachers' perceptual ability has been shown to decrease markedly with increasing age of pupils (3, 17).

SUMMARY AND CONCLUSIONS

Teachers' ratings of adjustment and aspirational traits were obtained for a class of fifty juniors in a University high school and related to various other self-report, objective, and projective measures of the same characteristics. The following conclusions can be reached with respect to our particular sample of adolescent boys and girls, and deserve to be tested on a more representative population:

- 1) Averages of five teachers' ratings of adjustment and aspirational traits enjoyed high split-half reliability.

- 2) Rorschach adjustment score, teachers' adjustment rating, and M.M.P.I. adjustment score were not significantly intercorrelated except for the latter two measures in the case of girls. Thus, all three instruments appear to be measuring quite different aspects of adjustment showing little or no overlap.

- 3) High level of anxiety, as measured by the Illinois Personality Inventory (I.P.I.) was consistently associated with poor adjustment in girls. This was not true of boys. The suggestion was offered that anxiety, especially the variety associated with high striving, is more compatible with and culturally acceptable as evidence of good adjustment in boys than in girls.

- 4) Teachers' ratings of adjustment appear to be more meaning-

ful and psychologically valid for girls than for boys in as much as they were correlated in the appropriate direction with M.M.P.I. and composite adjustment scores and with I.P.I. anxiety score. This sex difference is attributed to the prevailing cultural ambivalence regarding the adjustive value of anxiety for boys in the light of their social sex rôle. In the case of girls, teachers' ratings of adjustment also tended to be more consistently correlated in the negative direction with 'real life' measures of level of academic aspiration.

5) Teachers tended to give high adjustment ratings to pupils whom they perceived as persistent and scholastically competitive, who were highly accepted by their classmates, who had superior scholastic records, and whose academic aspirations were commensurate with past academic achievement.

6) Teachers' ratings of pupils' aspirational traits were very highly related to the latter's scholastic standing. Hence, academic performance can not be used as an independent criterion of the validity of these ratings.

7) In view of the reliably negative correlations obtained between 'real life' measures of level of academic aspiration and teachers' ratings of aspirational traits, the validity of the latter measure seems highly questionable.

REFERENCES

- 1) D. P. Ausubel. "Prestige motivation of gifted children." *Genet. Psychol. Monogr.*, 1951, 43, 53-117.
- 2) D. P. Ausubel and H. M. Schiff. "A level of aspiration approach to the measurement of goal tenacity." *J. gen. Psychol.*, (in press).
- 3) D. P. Ausubel, H. M. Schiff and E. B. Gasser. "A preliminary study of developmental trends in sociopathy: Accuracy of perception of own and others' sociometric status." *Child Developm.*, 1952, 23, 111-118.
- 4) D. P. Ausubel, H. M. Schiff and M. Goldman. "Qualitative characteristics in the learning process associated with anxiety." *J. abnorm. soc. Psychol.*, 1953, 48, 537-547.
- 5) D. P. Ausubel, H. M. Schiff and M. P. Zeleny. "'Real life' measures of level of academic and vocational aspiration in adolescents: Relation to laboratory measures and to adjustment." *Child Developm.* (in press).
- 6) H. L. Baker, "High-school teachers' knowledge of their pupils." *Sch. Rev.*, 1938, 46, 175-190.
- 7) M. E. Bonney, "Sociometric study of agreement between teacher judgments and student choices." *Sociometry*, 1947, 10, 133-146.
- 8) L. J. Cronbach, *Essentials of Psychological Testing*. New York: Harper, 1949.

9) R. M. Eichler, "Experimental stress and alleged Rorschach indices of anxiety." *J. abnorm. soc. Psychol.*, 1951, 46, 344-355.

10) Molly R. Harrower and M. E. Steiner, *Large Scale Rorschach Techniques: A Manual for the Group Rorschach and Multiple-choice Test*. Springfield, Ill.: Charles C Thomas, 1945.

11) S. R. Hathaway and J. C. McKinley, *The Minnesota Multiphasic Personality Inventory*. Minneapolis: Univer. of Minnesota Press, 1943.

12) E. L. Kelly and D. W. Fiske, *The Prediction of Performance in Clinical Psychology*. Ann Arbor, Michigan: Univer. of Michigan Press, 1951.

13) B. Klopfer and D. M. Kelly, *The Rorschach Technique*. New York: World Book Co., 1946.

14) G. F. Kuder, *Revised Manual for the Kuder Preference Record*. Chicago: Science Research Associates, 1946.

15) E. A. Lee and L. P. Thorpe, *Occupational Interest Inventory, Advanced Form A*. Los Angeles: California Test Bureau, 1946.

16) W. D. Lewis, "A comparative study of the personalities, interests and home backgrounds of gifted children of superior and inferior educational achievement." *J. genet. Psychol.*, 1941, 59, 207-218.

17) J. L. Moreno, *Who Shall Survive?* Washington, D. C.: Nervous and Mental Disease Publishing Co., 1934.

18) Ruth L. Munroe, "Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method." *Appl. Psychol. Monogr.*, 1945, No. 7.

19) D. Rapaport, *Diagnostic Psychological Testing*. Chicago: Yearbook Publishing Co., 1945.

20) P. M. Symonds, *Diagnosing Personality and Conduct*. New York: Appleton-Century, 1931.

21) L. M. Terman and M. Oden, *The Gifted Child Grows up: 25 Years' Follow-up of a Superior Group*. Stanford, Calif.: Stanford Univer. Press, 1949.

AN EXPERIMENT IN NINTH-GRADE READING IMPROVEMENT

ROBERT F. BARRY

Consultant in Research and Testing, Rochester, N. Y., Public Schools

PAUL E. SMITH

Assistant Superintendant, Rochester, N. Y., Public Schools

The Rochester public schools have recently experimented with several ways for improving reading ability. The experiments were designed for normal classroom use and were carried on with ninth-graders. Attention was centered particularly on methods which might be effective with poorer readers. This effort began in 1951-1952 with experimental use of the fourteen State University of Iowa Reading Films, and was extended in 1952-1953 to include a number of other group methods.

1951-1952 EXPERIMENT

Purpose.—During the school year 1951-1952 the State University of Iowa Reading Films were used experimentally in all ninth-grade classes of two high schools. The purpose of this experiment was to investigate the merit of the films in the improvement of reading ability.

In setting up the experiment, certain conditions had to be met: (1) the actual showing of the films had to be fitted into the regular schedule of classes; (2) the projection rooms could be used only in such a way as not to disrupt their use by other classes in the school; (3) practice reading material had to be readily available; (4) the test selected for measuring changes in reading ability had to have three parallel forms, each requiring not more than one class period for administering; (5) all parts of the experiment had to be such that they could be carried on by classroom English teachers who were not specially trained in the teaching of reading.

Procedures.—

Pre-test

1) Form A of the Nelson Silent Reading Test¹ was administered

¹ All results were based on total scores for the Nelson Silent Reading Test, which were composed of scores on the following separately timed sub-

to all ninth-grade pupils (four hundred and seventy-three) in two high schools in October 1951. The two schools selected were of average size, had heterogeneous populations, and were not atypical of other high schools in the city. This and all subsequent tests were centrally scored by experienced test clerks.

2) The fourteen State University of Iowa Reading Films were used with one half of the groups at the rate of one film a week, and with the other half at the rate of two films a week. Each film showing occupied three to five minutes at the beginning of the class.

3) After each film a ten-question objective-type test (supplied by the film producer) was administered in order to sharpen the attention of the pupils to the content of the film. Each pupil kept a record of his scores on these 'film tests'.

4) During the same class period the pupils read a selected *Reader's Digest* article and noted the length of time required for reading.

5) Each pupil converted his reading time into a reading rate for each article and maintained a continuing graph of his reading rates for the fourteen articles.

Mid-test

6) A few days after using the last film in the series, Form B of the Nelson Silent Reading test was administered. This was intended to measure the reading ability at the end of the film showings.

7) Then followed a 'retention' period of three and one-half months during which there was no particular classroom emphasis on reading improvement.

Post-test

8) After the expiration of this interval, Form C of the Nelson Silent Reading test was administered. This was intended to measure the reading ability three and one-half months after the showing of the films had been completed.

9) No control groups were established. Gains were measured in tests: vocabulary, reading for general significance, ability to note detail, and ability to predict probable outcomes.

TABLE I.—RESULTS OF PRE- AND POST-TESTING NINTH GRADES 1951-1952

Median Grade Equivalents		Median Gains	
At start of experiment (at 9.1)	At end of experiment (at 9.75)	Total	Beyond expectation
8.87	10.39	15 months	8.5 months

normal classroom situations, and were based upon the grade equivalents established by the test publisher.

Findings.—The findings for the ninth-graders as a group are shown in Table I.

1) There was no significant difference between the groups that saw one film a week and the groups that saw two films a week.

2) During the course of the experiment (six and one-half months) the median gain for the four hundred and seventy-three pupils was fifteen months, or a net gain of eight and one-half months beyond expectation (expectation meaning the number of months in the interval of the experiment).

3) After separating the pupils on a basis of their scores at the beginning of the experiment, (a) the median gain of the top quarter was five months beyond expectation, (b) the median gain of the middle half was eight and one-half months beyond expectation, (c) the median gain of the lowest quarter was fifteen months beyond expectation.

4) The distribution of the gains of the four hundred and seventy-three pupils showed that four hundred and forty-five pupils, or ninety-four per cent, gained more than expectation during the course of the experiment.

Questions Raised.—Consideration of the findings of the experiment brought up two significant questions:

1) Would the rather startling gains made during 1951-1952 be maintained, or increase still more, or decrease, during the subsequent year, if the pupils had no further experimentation with reading? This question might be answered by follow-up testing during the following school year.

2) Would it be possible to attain similar gains by using a simpler method made up of parts of the original method? This seemed particularly important because the original method involved certain difficulties, such as, (a) the problem of scheduling large num-

TABLE II.—COMPARISON OF GAINS FOR NINTH AND TENTH GRADE TESTING OF IDENTICAL PUPILS

Grade	Changes in Median Grade Equivalents	Interval	Comparative Gains
9th	14.5 months gain during 6.5 months (gr. level 8.95 to gr. level 10.4)	Oct. '51–May '52	8 months more than expectation
10th	4 months gain during 9 months (gr. level 10.4 to gr. level 10.8)	May '52–Apr. '53	5 months less than expectation

bers of classes with limited numbers of films and projection rooms, (b) eyestrain reported by some pupils, (c) frustration and confusion of some pupils, resulting in decreased motivation, (d) the artificiality of reading from a motion picture screen.

Follow-up.—This was an effort to determine the extent to which the gains made during 1951–1952 were carried on into the tenth grade when there was no particular emphasis on reading (Question 1 above).

The Nelson Silent Reading Test, Form C, was given in April to the tenth grade in the same two high schools. Table II includes three hundred and fifty-six identical pupils. The first line shows gains made during the ninth-grade experiment, the second line shows gains made during the tenth-grade when no experiment was being conducted.

1952–1953 EXPERIMENT

Purpose.—An attempt was made to answer Question 2; namely, would it be possible to attain similar gains in reading ability by using a simplified method?

Procedures.—1) All ninth-grade pupils in the Rochester public schools (2166) were pre-tested with the Nelson Silent Reading Test, Form A, in October, 1952. The tests were scored centrally by experienced test clerks.

2) The 2166 pupils were divided into the eight groups listed below, with a different method being used in each group. Each group contained some above-average, some average, and some below-average English classes. These eight methods for improving read-

ing ability were then used for the next seven weeks:

Method	Pupils and Schools	Description of the Method
A	284 pupils in four schools	14 timed articles from the <i>Reader's Digest</i> , graph of reading rate
B	148 pupils in two schools	14 reading films and film tests
C	253 pupils in four schools	7 reading films and film tests, 14 timed articles from the <i>Reader's Digest</i> , graph of reading rate
D	225 pupils in four schools	3 reading films and film tests, 7 timed articles from the <i>Reader's Digest</i>
E	121 pupils in two schools	14 reading films and film tests, timed articles from the <i>Reader's Digest</i> , graph of reading rate (same method as in first year)
F	229 pupils in two schools	14 timed articles from <i>SRA Better Reading Book II</i> , graph of reading rate
G	177 pupils in two schools	14 timed articles from <i>Let's Read</i> (Silver book), graph of reading rate
H	729 pupils in seven schools	No special treatment, pre- and post-testing only
Total	2166 pupils in nine schools	

3) This was followed by a 'retention' period of three and one-half months.

4) Pupils were then post-tested with the Nelson Silent Reading Test, Form C, in March, 1953.

5) No control groups were established. Gains were measured in normal classroom situations, and performances were expressed in terms of the grade equivalents established by the test publisher.

Findings.—1) Median gains for each of the eight methods are shown in Table III.

2) Considering the lowest quarter, the middle half, and the top quarter of the pupils, separated on the basis of their scores on the pre-test at the beginning of the experiment, the median net gains for each method (gains beyond expectation) are shown in Table IV.

3) It is worth noting in Table IV that these methods produced reasonable net gains with all groups of pupils and the largest net gains with the lowest quarter of pupils.

4) An examination of the data showed that there was apparently

TABLE III.—RESULTS OF PRE- AND POST-TESTING NINTH GRADES,
1952-1953

Method	Median Grade Equivalents		Median Gains	
	At start of experiment (at 9.1)	At end of experiment (at 9.6)	Total	Beyond expectation
A	9.2	10.3	11 months	6 months
B	8.9	10.0	11 months	6 months
C	9.0	10.2	12 months	7 months
D	9.0	10.2	12 months	7 months
E	8.4	9.7	13 months	8 months
F	9.3	10.5	12 months	7 months
G	8.6	10.0	14 months	9 months
H	8.7	9.9	12 months	7 months

no relation between the size of the net gains of the lowest quarter (Table IV) and the median of the lowest quarter at the start of the experiment in October, 1952. For example, the two groups with the lowest starting grade equivalents were E and H. These made neither the greatest nor the least net gains. Similarly, the two groups with the highest starting grade equivalents were D and F. They made neither the greatest nor the least net gains. The median starting grade equivalents ranged from 6.5 to 7.1.

5) With respect to the lowest quarter of the pupils, the difference from one method to the next is not large. It is worth noting, however, that the two groups that used the entire series of fourteen films (B and E, Table IV) made the smallest gains. The two groups

TABLE IV.—MEDIAN NET GAINS BY EACH METHOD FOR VARIOUS LEVELS
OF READING ABILITY NINTH GRADES 1952-1953

Method	<i>Lowest quarter</i> Below 7.8 gr. eq. in Oct. '52	<i>Middle half</i> 7.8 to 9.7 gr. eq. in Oct. '52	<i>Top quarter</i> 9.8 or higher gr. eq. in Oct. '52
A	14 months	6 months	2 months
B	7 months	7 months	3 months
C	12 months	7 months	4 months
D	13 months	7 months	4 months
E	9 months	8 months	7 months
F	11 months	8 months	6 months
G	18 months	7 months	7 months
H	11 months	7 months	4 months

that made the largest gains were the ones that only read some type of timed articles (G and A, Table IV).

6) The distribution of the gains of the 2166 pupils shows that 1842, or eighty-five per cent, gained at least as much as expectation during the course of the experiment.

CONCLUSIONS

1) All methods used, both in 1951-1952 and 1952-1953, were effective in raising reading levels of ninth-graders as a group beyond expectation.

2) Of the eight methods used in 1952-1953, there was no appreciable difference among the various methods in median net gains of the total groups.

3) Certain methods were appreciably more effective with the lowest quarter of the pupils.

4) In the 1951-1952 experiment startling gains in reading ability were associated with the State University of Iowa Reading films; however, in 1952-1953 similar gains were obtained without the use of reading films.

5) Pre-testing and post-testing alone appear to be a valuable method of motivation for reading improvement.

6) Gains more than expectation which occurred with experimentation changed to gains less than expectation when experimental procedures ceased.

IMPLICATIONS FOR CLASSROOM PROCEDURES

The experiments seem to imply that the critical factors in increasing reading ability of pupils are (1) the focusing of attention of the teacher on reading improvement of the pupils, and (2) the focusing of attention of each pupil on the possibility of his own self-improvement in reading. This may be accomplished through a systematic, planned procedure. It would appear that systematic regularity in the use of films and/or practice reading material is a strong motivating device. It is also apparent that pre-testing at the beginning of the school year, followed by post-testing several months later is a strong motivating force, providing that the testing itself is announced and explained in advance.

If the once-emphasized attention on reading improvement is removed, pupils generally will then fail to continue to improve at the increased rate. Review of the 1951-1952 and 1952-1953 experi-

ments indicates that these methods of improving reading result in a temporary acceleration of gains. When the emphasis is removed the gains tend to decelerate to normal or less, although the actual reading level (grade equivalent) may still be above the chronological grade level.

All the eight methods described in a preceding section were effective in improving reading ability of pupils, and particularly so for the poorer readers. Each method is relatively easy to use, and may be incorporated into most high-school classes with a minimum of interference with other classroom work. It would appear that classroom teachers might well consider any one of these methods as a means for reading improvement.

THE PSYCHOLOGY PREREQUISITE TO EDUCATIONAL PSYCHOLOGY

ROY GLADSTONE

Oklahoma Agricultural and Mechanical College

The question at which the investigations reported below were aimed was whether or not the competence of students in educational psychology at the Oklahoma Agricultural and Mechanical College is affected by the fact that they had previously taken or were concurrently taking a course in psychology. This purpose was tested by two slightly different designs.

DESIGN OF THE FIRST EXPERIMENT

Hypothesis: The hypothesis tested was that students who had previously taken some work in psychology would do no better in educational psychology than those who had taken none.

Criterion: The criterion of success used was ability exhibited on the various essay and objective tests submitted by the student. These were designed to test the ability of the student to apply psychological insights rather than to recognize or reproduce memorized material. A numerical error score was assigned to each piece of work handed in by the student. The error score on the objective tests was the number wrong. The error score on the essay materials was assigned by estimating how far short the student fell of the requisites stated in a model answer. The objective test errors contributed about 1.5 times as much to the total score as did the essay errors.

Subjects: The subjects used were drawn from the Fall, 1950, educational psychology classes of the writer. Two groups of subjects were defined. One of the groups was composed of subjects who had not had and were not concurrently taking any courses that were primarily psychological in nature. The other group was composed of individuals who had completed one or more courses which were primarily psychological in nature (whether taken in the department of psychology or not) before entering the course in educational psychology.

Matching: The next step was to draw matched subjects from the two groups. Matching took place in ignorance of the criterion

scores. The subjects were matched on three variables—ACE percentile, hours of 'outside' work, and sex. Twenty-eight pairs were matched. The mean difference between the two groups was 0.5 ACE percentile units. The mean absolute difference was 1.5 with a range of 0-4. The differences favored the group having had no psychology—a conservative error.

Twenty of the twenty-eight pairs were matched by sex.

In order to match the students on hours of work aside from school work, data were gathered by means of a check list consisting of five categories: 0-5 hours of work, 5-10, 10-15, 15-20, and more than 20. The students were instructed to consider as work anything which occupied their time which fell outside the "normal student living processes." They were instructed to consider housekeeping and varsity practice, but not studying, as work for this purpose. The matching on this variable was not as good as for the ACE (ACE scores were considered more important and given priority in the matching process), exhibiting a mean difference of 0.5 scale units in favor of those who had taken psychology, i.e., they worked about two and one-half hours less per week than those who had taken no psychology. This was a non-conservative error.

In all, it is considered that the matching was reasonably good.

Results: Of the 28 who had taken psychology, 20 did better than the subjects with whom they had been paired, 2 did just as well, and 6 did more poorly. Using a split of 20 to 8, this yields a chi-square of 5.14. With one degree of freedom this corresponds to a p-level of less than .05.

The mean difference between the two groups was 15.5 error units in favor of the group which had taken psychology. This corresponds to about one median deviation of the array of error scores.

EXPERIMENTAL DESIGN OF THE SECOND EXPERIMENT

Hypotheses: Three related hypotheses were tested. The first was that those students who had taken the first course only in a department of psychology would do no better in a course in educational psychology than those who had not taken psychology. The second hypothesis was that those who had taken the first course only in a department of psychology or one course only which was primarily psychological in nature, but given by a department other than psychology, would do no better than those who had no psychological background. The third was that those who had some psycholog-

ical background, however extensive, would do no better than those who had no psychology. These three phases will be referred to below as units 1, 2, and 3. The three units are cumulative—that is, unit 2 includes unit 1, and unit 3 includes both 1 and 2.

Criterion: The criterion of success was score on the standardized test, "A Study of Barry Black." This test is designed to test ability to apply psychological knowledge rather than to test memory. Its reliability is .79 (1).

Subjects: The subjects were drawn from the Fall, 1951, and Fall, 1952, educational psychology classes of the writer and the Fall, 1952, educational psychology class of another instructor. Again two groups of subjects were defined. One was composed of subjects with no psychology course background prior to the course in education psychology. The other was composed of subjects who had taken some psychology course either previously to or concurrently with the educational psychology course.

Matching: This time the subjects were matched on ACE score, hours of work, sex, college class, instructor, and year the course was taken. The matchings took place in ignorance of the scores made on the test. In no case was the student of one instructor matched with a student of the other. Neither was a student taking a course in one year matched with one taking a course in another year. The next most important variable was considered to be the ACE score and some accuracy in matching on the remaining variables was sacrificed in order to get a good match on the ACE score.

Table I indicates the effectiveness of the matching procedure. It will be noted that the average differences in ACE scores are all in favor of the group which had taken psychology, but the differences are so small that they were considered negligible. The differences in hours of work in favor of the 'with psychology' group are proportionately larger but still small. The differences in class standing were in favor of the 'no psychology' group. Matching on the variable of sex was not good but the significance of this is obscure. The mean absolute difference in ACE score was about half a standard error and in no case was a match made where the difference was greater than 7 or approximately one standard error. In all it is considered that the matchings were adequate. Matchings for unit 2 were best.

¹ I am indebted to Dr. Roy Sommerfeld of Oklahoma A. and M. College for his coöperation.

TABLE I.—MATCHING FOR THE THREE UNITS

	Unit 1	Unit 2	Unit 3
Number of pairs	28	36	39
ACE mean algebraic difference ^{a, b, c}	-0.32	-0.08	-0.13
ACE mean absolute differences ^b	2.9	2.9	2.9
Range of ACE differences ^b	7	7	7
Work mean differences ^{c, d}	-0.25	-0.06	-0.15
Class mean difference ^{c, e}	0.25	0.25	0.21
Number of pairs matched on sex	20	21	23

^a The standard deviation of the ACE combined raw score is about 25 (2).

^b The standard error of the ACE total raw score is of the order of 7 (2).

^c A minus indicates a difference which is presumed to be in favor of the group which had taken psychology—a non-conservative error.

^d Five-hour classes make up this scale. Thus a difference of 0.25 indicates about one and one-quarter hours per week difference in time spent working.

^e The freshman class was assigned a value of one; the sophomore, two, etc.

TABLE II.—FINDINGS OF THE SECOND EXPERIMENT²

	Unit 1	Unit 2	Unit 3
Average difference between means ^{a, b}	10.93	22.03	18.36
t-scores of difference	1.07	2.29	1.95

^a The difference in raw score between the 40th and 60th percentiles is 20; that between the 80th and 90th is 20 (1).

^b All three differences are in favor of the 'with psychology' groups.

Findings: The results of the second experiment are summarized in Table II.

If the one-tail test is used, all three t-scores correspond to the five per cent level or less.³ If the two-tail test is used only the second unit reaches the five per cent level of significance, that for the third unit falling slightly short of that level ($t = 2.025$ for 38 degrees of freedom at the five per cent level).

² I am indebted to the Statistical Laboratory of Oklahoma A. and M. College for the calculation of the t-scores.

³ The use of the one-tail test seems justified since the experiment was designed to test only whether having had psychology makes a student better able to cope with educational psychology. There is no reason to suspect that a complete reversal of this proposition would occur.

SUMMARY

Two experiments were designed to test the efficacy of a psychological background as an aid to the student in educational psychology. Both experiments were paired experiments and differed essentially only in that the criterion of success in the first was a total score on all the teacher-made assignments of the course in educational psychology while the criterion of success in the second was the score on a standardized test.

In the first experiment the difference in favor of the 'with psychology' group was fairly large and the five per cent level was reached through the use of the fairly crude chi-square. The second experiment was analyzed in three stages. All three showed a difference in favor of the 'with psychology' group. Using a paired analysis and the one-tail test all three stages were significant at the five per cent level. Using the two-tail test only the second stage was significant at the five per cent level.

DISCUSSION

It is interesting to speculate concerning the reasons why the difference in the second experiment was smaller than that in the first. One possible reason is the fact that, having analyzed the results of the 1950 experiment, the instructor told the two successive classes that those without the prerequisite psychology were at a disadvantage and gave the class some material designed to help the student bridge the gap. Another possible reason is that the criterion of the first experiment was both longer and broader—there were more objective questions and some essay material was included.

CONCLUSIONS

It is concluded that a background of psychology will help the students make a higher grade in the course in educational psychology at Oklahoma A. and M. College. Whether or not this would be true of other colleges could easily be a function of the nature of courses which are given. The differences actually obtained were of sufficient magnitude to affect the grade by one grade point, i.e., a student without a background of psychology who received a C would likely have received a B had he had a course in psychology. While this is hardly world-shattering, it is certainly of sufficient importance to be taken into account.

Lastly, it should be explicitly stated that there is no implication in this material that a prerequisite course in psychology affects only the grade in educational psychology. It is reasonable to assume, for instance, that class discussions can go faster and further when the students have some psychological background. Investigations of such factors would be very desirable.

REFERENCES

- 1) John E. Horrocks and Maurice E. Troyer. *A Study of Barry Black and Instructor's Guide and Norms*. Syracuse, N. Y.: Syracuse University Press, 1948.
- 2) American Council on Education. *Psychological Examination for College Freshmen and Manual of Instructions*. Princeton, N. J.: Educational Testing Service, 1948.

AN EXPERIMENTAL STUDY OF THE READING-RATE CONTROLLER

GEORGE F. WOOSTER¹

Ohio State University

For a number of years the use of various mechanical devices, such as the reading-rate controller, the metronoscope, the Harvard Reading Films, and the tachistoscope, for improving reading has been one of the most controversial subjects in the field of reading. Opinions on the topic range from completely favorable to wholly unfavorable. A large percentage of the studies published with regard to this subject deal with the use of the Harvard Reading Films and the metronoscope. While commercial representations concerning reading-rate controllers have increased, there has been practically nothing in the professional literature about the effectiveness of their use.

Most of the reports on the use of these devices (other than the reading-rate controller) have been quite favorable. They tend to indicate that fine results may be expected. Many of these studies have been well done; there is one aspect, however, which should be noted—in each case, experimental groups were compared only with control groups which had been given no training at all.

A report from Lauer, Henry, and Fritz (5), and one from Henry and Lauer (4), deviate from this pattern. These people, working with college students, compared four methods of improving reading, as well as comparing the experimental groups with a control group having no practice. (A mechanical device, the metronoscope, was used with only one of their experimental groups.) All their experimental groups showed definite improvement, with no one group significantly better than the others. These investigators concluded that any reasonable technique involving the reading function would result in improvement, and artificial devices may be more spectacular than useful.

In view of the claims made concerning the advantages of the reading-rate controller as a means for improving reading, more

¹ This study was the basis for a M.A. thesis done in the Department of Psychology at Ohio State University under the direction of Dr. John R. Kinzer.

data are needed. The study here considered, although carried out on a small scale, was set up in an attempt to gain some data of the type needed with regard to these more recent mechanical devices. We were directly concerned with the possible value of supplementary training on the reading-rate controller in conjunction with a course in effective study.

Reading is only one of several topics considered in this effective study course. (See Francis P. Robinson, *Effective Study*; this book is used as a basic workbook for the course.) The reading process and ways of improving reading are discussed in class after the students have read the project in *Effective Study*. Generally, three or four class periods (or portions of several periods) are set aside during the quarter for practicing speeded reading. Further consideration is then given to reading as it may be a part of other topics covered, such as the SQ3R (Survey, Question, Read, Recite, Review) method of study.

If students wish to work on improving their reading skills in addition to what is done in class, individual help may be obtained from the instructor. Some students may also obtain individual help from a graduate counselor. (Practicum training of counselors includes a period of counseling with students who are taking the Psychology 411 course.) In such cases, there is a reading-rate controller available for individual use if desired, as well as different types of practice reading material.

While improvement of reading is not, in itself, the major purpose of this effective study course, experience has shown that during a quarter (approximately ten weeks) a large percentage of the students can be expected to show some improvement. Changes in reading rate and comprehension are indicated by a comparison of test scores, particularly those obtained on two parallel Robinson-Hall Reading Tests which are given at the beginning and the end of the quarter. Knowing that some improvement could be expected, the question to be considered was whether or not greater improvement would be achieved if students were given training on the reading-rate controller in addition to the training received in the course.

PROCEDURE

All students taking Psychology 411 (Psychology of Effective Study and Individual Adjustment) during the spring quarter of

1950 were tested on the Robinson-Hall "Reading Test in History: Canada" during the first week of the quarter. After this, two groups of six students each were selected, on recommendation by the instructors of the several classes, from the large number who volunteered for the project. Lack of available facilities limited the size of the groups.

An attempt was made to set up matched groups. The problem of scheduling training periods for each of the students, however, and the difficulty encountered in trying to match the groups on the basis of the reading test score, made it evident that any matching could only be approximate. Consequently, for statistical purposes, the groups were treated as unmatched groups.

The training sessions were scheduled during a four-week period. Experimental Group A received three hundred forty-five minutes of training on the reading-rate controller in eleven half-hour sessions and one fifteen-minute session. Experimental Group B received three hundred fifteen minutes of training on the controller in ten half-hour sessions and one fifteen-minute session.

The device used was the reading-rate controller manufactured by the Three Dimension Company of Chicago, Illinois. (A description of the apparatus, the mechanics of operation, and the manner of preparing material can be found in the manual, *Reading-Rate Controller*, by Helen M. Robinson, Director of the University of Chicago Reading Clinic.) The reading material used consisted of book sections taken from various issues of the *Reader's Digest* and some Public Affairs Pamphlets.

Each half-hour session was divided into two reading periods of fifteen minutes each, with a five-minute interval between the two periods of reading. This allowed for a change of the rate setting on the controller within a session as well as from session to session. Times for increasing the rate setting were determined on the basis of the subject's report that he could keep up with the machine, or ahead of it, without difficulty, and an oral comprehension check. (The rate was increased only if the comprehension level as checked remained at seventy per cent or better.) Increments of increase in the rate setting were kept small—ten to fifteen words per minute. This allowed for more frequent increases, which it was believed would serve as a further motivating factor, and aided in keeping a running check on comprehension. Subjects were not told at what rate they were reading unless they inquired about it.

At the end of the training period, the twelve subjects, as well as all other students taking the course, were tested on the "Reading Test in History: Russia," a test form comparable to the pre-training test.² The scores from this test could then be compared with the scores obtained on the "Reading Test in History: Canada," the pre-training test, which provided a basis for comparisons both within the groups and between the groups. Scores on both these tests were available for a group of forty-eight students in these classes in addition to the twelve subjects in the two experimental groups. The former group was then considered as a control group to determine if those who had received training on the rate controller in addition to taking the effective-study course showed significantly greater changes in reading ability than those who took the course but had no supplementary training on the reading-rate controller.

RESULTS

Comparable pre-training and post-training scores on rate and comprehension were obtained by using two parallel reading tests from the Robinson-Hall series. With these scores, a comparison was made for the subjects within each group, and the means of the three groups were compared with each other. All differences were tested by the use of the t-test.

From an examination of the data thus obtained, the following conclusions may be drawn:

A) With regard to rate:

- 1) The gain in rate was significant at the two per cent level for the two experimental groups, and at the one per cent level for the control group.
- 2) Variability was greater on the posttest; the standard deviation was greater for the scores on the posttest in each group.
- 3) Decreases in rate were evident for only eight students, all of whom were included in the control group.
- 4) Comparing the mean gains in rate among the three groups, small differences were noted. None of these differences, however, was statistically significant.

² For information on the comparability of these two tests, see the *Manual for the Robinson-Hall Reading Tests*, Ohio State University.

TABLE 1.—THIS TABLE SHOWS THE GROUP MEANS ON THE PRETEST AND POSTTEST FOR RATE AND COMPREHENSION

Group	Mean Test Scores			
	Rate (words per minute)		Comprehension (per cent right)	
	Pretest	Posttest	Pretest	Posttest
Experimental—A	189.83	237.83	69.33	63.50
Experimental—B	184.83	220.33	81.33	73.67
Control group	181.80	226.40	64.19	65.17

B) With regard to comprehension:

1) In each experimental group, one-half or more showed some decrease in score; nearly one-half of the control group also showed some decrease.

2) Variability within the group was less for the posttest than for the pretest; this was true in all three groups.

3) No clear pattern of relationship between amount of gain or loss in rate and amount of gain or loss in comprehension was noticeable.

SUMMARY

This brief study has been concerned with the possible value of supplementary training on the reading-rate controller in conjunction with a course in effective study. As measured by two parallel forms of the Robinson-Hall Reading Tests in History, most of the students taking the course showed some improvement in rate of reading. As a group, the mean difference in rate between the pretest and posttest was statistically significant ($P = .001$) for those who did not have additional training. There was no statistically significant difference in comprehension scores.

Students receiving additional training on the reading-rate controller also showed significant improvement ($P = .02$) in rate of reading. Again, there was no statistically significant difference in comprehension scores. Gains made by these experimental groups, however, were not significantly different from the gains of the control group.

In this case, no extra benefits seemed to accrue from additional training on the reading-rate controller, as used. The study sheds

some light on the possible value of using such a device in connection with a course where reading is treated only as one of several topics. Further investigation is needed, however, concerning the practical usefulness of these machines and the comparative effectiveness of different approaches to the problem of improving reading skills.

REFERENCES

- 1) R. B. Ammons and A. M. Hieronymus. "Critical evaluation of a college program for reading improvement." *Journal of Educational Psychology*, Vol. 38, 1947, pp. 449-470.
- 2) Guy T. Buswell. *Remedial Reading at the College and Adult Levels*. Supplementary Educational Monograph No. 50. Chicago: University of Chicago, Nov., 1939, p. 69.
- 3) W. F. Dearborn and S. V. Wilking. "Improving the reading of college freshmen," *School Review*, Vol. 49, Nov., 1941, pp. 668-678.
- 4) L. K. Henry and A. R. Lauer. "A comparison of four methods of increasing the reading speed of college students," *Iowa Academy of Science*, Vol. 46, 1939, pp. 273-276.
- 5) A. R. Lauer, L. K. Henry and J. F. Fritz. "Improving the reading ability of college students," *Psychological Bulletin*, Vol. 35, 1938, p. 670.
- 6) George W. Snedecor. *Statistical Methods*. Ames, Iowa: The Collegiate Press, 1946.
- 7) A. E. Traxler. "Value of controlled reading: Summary of opinion and research," *Journal of Experimental Education*, June, 1943, pp. 280-292.
- 8) V. E. Witzeman. "An experimental study using the Ophthalm-O-Graph and Metron-O-Scope in the diagnosis and treatment of reading defects," *The Journal of Psychology*, Vol. 11, 1941, pp. 307-334.

RELATIONSHIPS BETWEEN EIGHT VARIABLES AND F-TEST SCORES OF TEACHERS

H. H. REMMERS AND MARTIN STEINBERG

Purdue University

Many investigators have concerned themselves with the problem of the organization of attitudes. Hartshorne and May (1), for example, have concluded that it is unjustifiable to consider honesty, service, and self-control as general characteristics of children, but rather that these traits are a function of the situation. On the other hand, Herrick (2) has drawn conclusions in favor of general attitudes of college students concerning social issues and matters of conduct. Gough (3) has found that the more antisemitic subjects in his samples are also characterized by lower intelligence, disadvantaged economic background, conservatism, and less sociability, to name but a few of the relationships discovered. Adorno and others (5) came to similar conclusions in their extensive study.

The present investigation is tangential to those mentioned above. The major interest here is in the factors related to certain attitudes. Struening (4), with university faculty members as subjects in a study similar to the present one, found a correlation of $-.45$ between attitude toward the church and F-test scores. Since low scores on the measure of attitude toward the church indicate favorable attitudes, the correlation indicates that the more favorable individuals on this continuum tend to be more authoritarian. In addition, Struening found that a favorable attitude toward the church was positively related to church membership. Another interesting finding was that the most favorable group was Catholics, then Protestants, and at the least favorable end of the continuum, Jews. This same order prevailed in F-Test scores, with Catholics at the more authoritarian end of the continuum. An attempt is made in this study to determine the relationships, if any, between scores of elementary- and secondary-school teachers on the Social Attitude Scale of Adorno et al (5)

and nine other variables. The ten variables are as follows:

- 1) Scores on the Social Attitude Scale or F-test or SAS (5)
- 2) Scores on the Sim's Social Class Identification Scale, or SCI (6)
- 3) Attitude toward the church, as measured by the Kelley-Remmers Scale of Attitude toward Any Social Institution, or SAI (7)
- 4) Sex (weighted 0 for male, 1 for female)
- 5) Age
- 6) Church attendance (weighted 0, 1, 2, and 3 for increasing frequency of attendance)
- 7) Home discipline (weighted 0 and 1 for lenient and strict, respectively)
- 8) Political party preference, or Pol. (weighted 0, 1, and 2, for Republican, Democrat, and Independent, respectively. An assumption is made here for correlational purposes that a continuum is represented by the order of code numbers.)
- 9) Race (weighted 0 for white, 1 for Negro)
- 10) Religion (Protestant, Catholic, or Jewish)

COLLECTION AND ANALYSIS OF DATA

The data were collected from one hundred eighty graduate students in education at the University of Southern California in July 1952. Practically all of the members of the group were teachers or school administrators. All intercorrelations among the variables were computed, and three multiple correlations with F-test score as the dependent variable were calculated. The three multiple correlations were for Protestants, Catholics, and for the total sample, including Protestants, Catholics, and Jews. A separate multiple correlation for the Jewish sample was ruled out due to the small sample size ($N = 8$).

INTERPRETATIONS AND CONCLUSIONS

1) It must be remembered that the correlations presented are probably underestimates of the population parameters, since the group was extremely homogeneous. As concerns the curvilinearity of the church attendance variable, Table 4 indicates that none of the η 's were significantly greater than the original correlation between the same two variables.

2) Although the mean scores of Catholics do not differ significantly from those of Protestants on any variable, and despite the fact that the correlation matrices for these two groups are very similar, still the multiple correlation for the Catholic sample is significantly greater than that for the Protestant group at far beyond the one per cent level.

3) An eta of .248 between frequency of church attendance and SAS scores for the total sample is found in Table 1. In addition, the eta representing the relationship between frequency of church attendance and attitude toward the church is found to be .290. These etas are both statistically significant, at better than the one per cent confidence level. This investigation, then, has substantiated Struening's findings, mentioned above, with regard to the correlation of church attendance with both attitude toward the church and authoritarianism scores. In Table 1 the correlation between F-test scores and attitude toward the church is also found to be significantly large.

TABLE 1.—CORRELATION MATRIX FOR TOTAL SAMPLE

	SCI	SAI	Sex	Age	Church	Home	Pol.	Race
Soc. Att. Scale	.024	.161†	.044	.003	.193*	.122	-.174†	.168†
Soc. Class Inv.		.226*	.263*	-.220*	.004	.051	-.170†	-.119
Att. to Church			.142	-.038	.281*	.011	-.033	.031
Sex				.288*	.172†	.199*	.001	.080
Age					.006	.132	.053	.060
Church Attend.						.120	-.173†	.180†
Home Discip.							-.055	.052
Polit. Pref.								.122

* Significant at the 1 per cent level.

† Significant at the five per cent level.

TABLE 2.—MULTIPLE CORRELATIONS OF ALL VARIABLES WITH SAS

	R 1.23456789	ST
Protestants (N = 145)	.323*	.075
Catholics (N = 27)	.694*	.102
Total (N = 180)	.330*	.067

* Significant at the one per cent level.

TABLE 3.—MEANS OF ALL VARIABLES AND STANDARD DEVIATIONS OF CONTINUOUS VARIABLES

	SAS	SCI	SAI	Age	Sex	Church	Home	Pol.	Race
Protestant (N = 145)									
Mean	-16.85	19.54	9.33	34.06	.26	1.26	.61	.70	.04
SD	21.54	4.83	1.77	7.14					
Catholic (N = 27)									
Mean	-12.93	19.85	9.44	33.63	.37	2.07	.63	.59	.15
SD	21.65	5.42	1.86	6.81					
Total (N = 180)									
Mean	-17.09*	19.56	9.26	33.86	.26	1.34	.59	.72	.06
SD	22.08	4.91	1.85	7.04					

* Jewish group pulled mean for total sample in a negative direction.

TABLE 4.—SIGNIFICANCE OF DIFFERENCES BETWEEN LINEAR CORRELATIONS AND ETAS FOR CHURCH ATTENDANCE VS. FOUR NON-CODED VARIABLES

	Linear r	Eta	P
Protestants (N = 145)			
Church vs. SAS	.111	.209	>.05
Church vs. SAI	.287	.289	>.50
Church vs. SCI	.025	.187	>.05
Church vs. AGE	-.023	-.107	>.30
Catholics (N = 27)			
Church vs. SAS	.437	.477	>.30
Church vs. SAI	.099	.215	>.50
Church vs. SCI	-.229	-.251	>.50
Church vs. AGE	.087	.136	>.50
Total (N = 180)			
Church vs. SAS	.193	.248	>.05
Church vs. SAI	.281	.290	>.50
Church vs. SCI	.004	.146	>.10
Church vs. AGE	.006	.079	>.50

Other evidence derived from the statistical analysis points in the same direction. Regression equations for Catholics and Protestants were computed from the correlation matrices. It was found that the most heavily weighted factor for predicting F-test scores for the Catholic sample was frequency of church attendance. The most heavily weighted factor for the Protestant sample was race. Attitude toward the church and frequency of church attendance ranked fourth and seventh, respectively, in this regard. This is indirect substantiation of Struening's finding that Catholics ranked higher than Protestants in scores on the F-test.

4) There are three possible explanations for these findings. First, persons who are authoritarian may find that some of their needs are satisfied by affiliation with the church. The second explanation might be that the education one receives through the medium of the church is authoritarian in character, and that this education carries over into other areas of life. Third, there may be an independent third factor which influences both authoritarianism scores and attitudes toward or membership in the church. Further research needs to be directed at these alternative hypotheses.

REFERENCES

- 1) H. Hartshorne and M. May. *Studies in the Nature of Character*, New York, The Macmillan Company, Vol. I-III, 1928-1930.
- 2) V. E. Herrick. "The generality and specificity of Attitudes," Ph.D. thesis, University of Wisconsin Library, unpublished; referred to in Kimbell Young, *Personality and Problem of Adjustment*, New York, F. S. Cross and Company, 1940, p. 288.
- 3) Harrison G. Gough. "Studies of social intolerance," *Journal of Social Psychology*, 1951, 33, 237-269.
- 4) Elmer Struening and Bernard Spilka. "A study of certain social and religious attitudes of university faculty members." *Psychol. Newsltr.*, 1952, No. 43, 1-18.
- 5) T. W. Adorno and Others. *The Authoritarian Personality*, New York, Harper and Brothers, 1950, 255-257.
- 6) Verner M. Sims. "A technique for measuring social class identification." *Educational and Psychological Measurement*, 1951, 11: 541-548.
- 7) Ida B. Kelley and H. H. Remmers. *A Scale for Measuring Attitude toward any Institution*, Forms A and B. Division of Educational Reference, Purdue University, 1934.

HOSTILITY: AN EDUCATIONAL PARADOX

HARRY A. GRACE

Michigan State College¹

Investigators of hostility have been primarily concerned with prejudiced people, anti-minority feelings, or ethnocentrism. This bias of investigation has precluded the possibility of analyzing hostility *per se*, devoid of special targets. We are concerned with the pattern of hostility', regardless of the particular target. Once we construct a systematic theory of hostility clarification of target specificity will readily follow. We are not interested in which target is most or least liked, but, given the rank order of any group of targets, how hostility is patterned.

We postulate that action is dependent upon intellectual and emotional forces. Rather than try to test such a broad thesis, we concentrate upon the hostile act as dependent upon both information and attitude. Should this be systematically demonstrated, social behavior theory may be anchored in a general theory of behavior. This paper presents data which cross-validate our hypothesis.

Instruments were developed to measure subjects' verbal reactions to various catastrophes in terms of the amount of information they felt they had about them, their attitude toward them, and the amount they felt threatened by them. Empirical results of hostile attitudes toward nations indicated that one has to feel one has some knowledge about objects in order to be hostile toward them. (1)

METHOD

The instruments prepared for this study offer a method of test construction which is easily administered and has high reliability. Modified paired comparison and Likert scaling techniques are combined to yield a flexible instrument which measures both direction and intensity of attitude. The test may be machine scored. It may be applied to preference inventories, attitude scales, or sociometrics.

¹ Milton R. Canfield, Iowa State College, and James F. Guyot, Michigan State College, cooperated on this study.

Subjects were sixty undergraduate students at the University of Illinois and Michigan State College.

Eleven catastrophes were selected on the basis of their diversity, likelihood of occurrence, familiarity, etc. They were: anemia, cancer, diabetes, heart attack, leprosy, nervous breakdown, pneumonia, poliomyelitis, syphilis, traffic accident, and tuberculosis.

For every pair in each inventory the subject marked a preference for one catastrophe (or indifference) to complete one of the following statements:

I have greater dislike for . . . than . . .

I have more accurate knowledge and information about . . . than . . .

I feel more threatened by and fearful of . . . than . . .

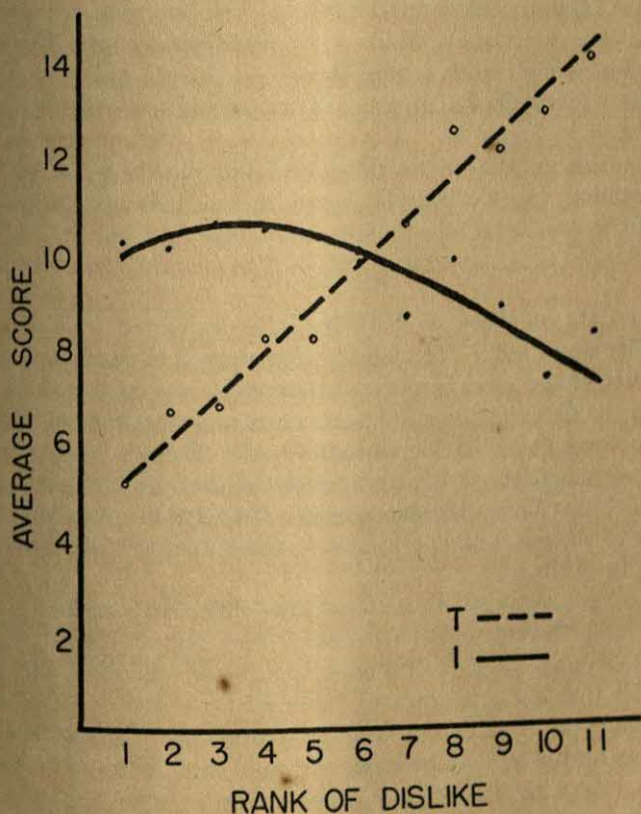


FIGURE 1.—Average Scores for Threat (T) and Assumed Information (I) According to the Rank Order of Dislike.

An item might look like this:

ANEMIA more about equal more CANCER

The subject marked his choice on an IBM answer sheet. Each catastrophe was scored separately. Two points were given for every 'more' answer and one point for an 'about equal' answer.

RESULTS

The intensity scores for each stimulus were plotted on the corresponding ranks to check linearity. The dislike and information stimuli were both linear, but the threat was curvilinear.

Stimuli were ranked in order of dislike for each subject. Rank one might include several stimuli, for different subjects selected various catastrophes as most disliked. The information scores for the attitude stimuli classified on the basis of ranks of dislike were determined and plotted. The hostility (threat) scores for these stimuli were similarly determined and plotted. The smooth curves are shown in Figure 1.

The data indicate a curvilinear relationship between assumed information and dislike. Threat is directly, linearly related to dislike.

DISCUSSION

The data on threat and dislike differ somewhat from data on hostility and liking. Curvilinearity appears in both cases, but the direction of the curve is reversed. This difference may be attributable to the different items (nations or catastrophes), or to different instructions (hostility or threat, like or dislike). We favor the latter explanation. It is corroborated by the sociometric analyses in which one finds differences in the distributions of positive and negative choices. Liking may not be the antonym of disliking, nor hostility of threat.

Another explanation is that these differences represent two patterns of hostility. Let us assume they are. Then, if it is our task to educate persons with regard to a set of stimuli, it is entirely possible that our program may have serious effects upon hostility. The more information one feels he receives about things he likes, the less hostile he may become; the more that he thinks he knows about things he dislikes, the more hostile he may become. Conversely, we feel we have greater information about stimuli which

we consider moderately threatening to us, those which we feel do not threaten us and those which threaten us severely we feel less informed about. In either case the effect of additional information will not be linear.

This may explain some of the boomerang effects in our educational efforts. Information behaves like so much ammunition when our dislikes are strong. When we are under threat we may not even perceive additional information. Before we can predict the outcome of a program of hostility or threat reduction we must first ascertain the pattern upon which our efforts of change ~~can~~ be exerted. Otherwise we are apt to increase the feelings of hostility and not affect threat in the least. The paradigm of hostility has a wide range of basic and applied research possibilities and offers a solution to a practical paradox in education. Guiding hypotheses for future research are:

1) Hostility (threat) is dependent upon the relationship between an individual's attitude toward a class of stimuli and the amount of information he feels he has about it.

2) The pattern of hostility may be changed by altering either attitude or amount of information, or both of these variables.

3) The ease and significance of changing a pattern of hostility depend upon the pattern held and its generality to other stimuli.

SUMMARY

The genesis of a systematic theory of hostile behavior is presented. Hostility is considered as the resultant of attitude and assumed information. A method for the study of hostility patterns is described. Results indicate that the relationship between attitude and assumed information is curvilinear. The educational implications of this curvilinearity are discussed.

REFERENCE

- 1) Harry A. Grace and Jack Olin Neuhaus. "Information and social distance as predictors of hostility toward nations," *J. abnorm. soc. Psychol.*, 1952, 47, 540-545.

BOOK REVIEWS

JEROME M. SEIDMAN. Editor. *The Adolescent: A Book of Readings*. New York: The Dryden Press, 1953, pp. 798. \$4.50.

Sixty-seven selections, in six parts and twenty-two 'chapters' may or may not be worth four dollars-and-a-half of an undergraduate's money. Aside from its binding and paper (which in this case are below average) the value of a book lies in its content. In an anthology the content depends upon the care and wisdom with which the editor has sampled the field, and upon the degree of integration he is able to achieve with his comments and organization. Finally, the sample of content depends upon the population of content.

Perhaps it is this last point which is the key to Seidman's *The Adolescent: A Book of Readings*. Every discipline which deals with human beings might touch upon the particular class of them which is called 'adolescents'. In fact, Dr. Seidman has included—as a kind of horrible sample—one short piece which is from no scholarly discipline at all, but which belongs, rather, to the profession of journalism. It is called "How To Be a Woman," and "although the reader may derive some insight from it, he will, of course, recognize that this article is written on a popular rather than a scientific level." To criticize adversely on the basis of these three pages would be unfair except that they illustrate what seems to this reviewer an excessively catholic taste. Anthropometry, anthropology, education, physiology, psychoanalysis, psychology and sociology are all represented. The selections range from the severely empirical and methodological paper of Gough and Peterson ("The Identification and Measurement of Predispositional Factors in Crime and Delinquency") through a labored excerpt from Lewin that strains to describe the problems of this period in field-theoretical terms, to the provocative but very interpretive observations of Margaret Mead and Caroline Zachry. In short: the sample is too broad and too shallow. Too broad, because most undergraduates will be unable to put many of the unrelated selections into any perspective at all. Too shallow, because so much breadth has been attempted that not enough information about any single problem can be included.

Some of this difficulty results from a population of writings

defined in terms of subjects rather than subject matter. The implication is that adolescents are a breed apart, which is doubtful. On the other hand their parents may well agree, and in any event the catalogues of many colleges and universities make this concession to student interest. But there is then no reason to include materials which deal with quite different topics and ages. Among these are Oscar Lewis' "Husbands and Wives in a Mexican Village," Gordon Allport's "The Resolution of Intergroup Tensions," and Murray and MacKinnon's "Assessment of Osgood's Book."

To the foregoing general, or principled, objections must be added another which is the occupational hazard of the anthologist: why one particular choice and the omission of another? Individual differences of opinion of this sort combined with the fundamental difficulty in the scope of the collection will cause many an instructor to think long and hard before asking his students to buy it. No instructor would assign all of it, but any instructor will find at least some things he would like all of his students to read. Whether many will find enough to justify the addition to their students' financial load is hard to say.

JOHN P. MCKEE

University of California

CHARLES E. OSGOOD. *Method and Theory in Experimental Psychology*. New York: Oxford University Press, 1953, pp. 800. \$10.00.

Following the large Handbook of Murchison (1934) and the much more recent volume edited by Stevens, we have here another monumental and detailed discussion of experimental psychology. The author states that his principal purpose is to provide a "text that evaluates experimental literature in close relation to critical theoretical issues." The attempt has been "to provide an interpretive framework that both brings out the significance of existing research and directs the reader's thinking toward as yet unsolved problems that need further research." It is presented both as a text and as a book of reference. "Although the coverage is extensive, it is also selective." To overcome "the potential disadvantage of single authorship," the author submitted many chapters "to one or more experimentalists who are experts in the fields concerned." (v).

There are four parts into which the subject matter has been divided. Part 1, Sensory Processes, includes, sensory quality, sensory intensity, audition and vision; Part 2, Perceptual Processes, includes, the nature of perceptual organization, projection dynamics in perception, and central dynamics in perception; Part 3, Learning, deals with fundamental operations in learning, theories of learning, certain controversial issues in learning theory, neurophysiology of learning, serial and transfer phenomena, and retention; Part 4, Symbolic Processes, discusses problem-solving and insight, thinking, and language behavior.

According to the preface this work is the result of about ten years of effort given primarily to this project. It has the unity that the thinking of a single individual can give to it. If one looks for the report of a certain experiment it may not be found. But on the whole, critical judgment must recognize the magnitude of the contribution in which so much of experimental results and theoretical considerations have been collected and interwoven into a well-organized presentation. Stress was made upon those studies that seemed to be most pertinent to the theoretical issues. In general the selection has been excellent and the advanced student of psychology will find a thoroughly valuable contribution to the understanding of the subject in its present stage of development.

The critical student will perhaps be helped to see the generally unsatisfactory nature of theoretical explanations as they exist today, and to realize the need for much more fundamental research for the best progress of the science. Not only this, but he may find definite suggestions for the planning of much needed research.

One of the valuable contributions of the volume lies in the inclusion of certain material not before brought together in a single book, such as synaesthesia, certain material in connection with problem-solving and transfer, and in relation to language behavior and communication.

Students of educational psychology should find valuable contribution to their thinking especially in connection with the discussions of learning, transfer, problem-solving, thinking and language behavior. They might also be helped to recognize the fact that theoretical explanation based on insufficient experimental data has been inadequate in psychology and too often carried over into unsatisfactory educational theory. And, perhaps again, they

may be helped to see that theory which conflicts with physiological or neurological facts already known, is to be very carefully avoided.

The book will be found to be especially valuable for reference work. The bibliography and author index contains 1290 citations. The subject index is well made and extensive, pages 781-800. A large number of figures and graphs add to the interest and clarity of the presentation.

This is one of the decidedly valuable contributions to psychology. It is scholarly, generally clear in statement, and not too many complicated and difficult definitions. It may be expected for a long time to be an indispensable book especially for advanced students and as a reference for both student and instructor. For any who may be inclined to despair because of the various isms and ologies within psychology, this contribution may help them to see that psychology has been developing out of the 'psychologies' and that, as in other sciences, the good will be sifted out from the bad, so that a sound psychology will be established and will prevail over the part-psychologies and the pseudo-psychologies that may have been inevitable outbreaks in the development of the science.

The author and publisher are to be congratulated.

A. S. EDWARDS

The University of Georgia

ROBERT LINDNER. *Prescription For Rebellion*. New York: Rinehart and Company, Inc., 1952, pp. 305. \$3.50.

If to improve mankind is our goal, what the world needs is to abandon the fallacy of adjustment and cultivate instead an appreciation for the indispensability of its alternative—positive rebellion as a pattern and way of life. This, in essence, is the theme of Lindner's *Prescription for Rebellion*. Admittedly, the volume is designed to sound an alarm. The content of the volume he presents under two sections: Book One, called Diagnosis, is composed of six chapters captioned as follows: The Lie that Binds; Shamans and Shibboleths; The Eleventh Commandment; Rebel's Fate; Adjustment—Causes and Consequences (1) History from a New Dimension; Adjustment—Causes and Consequences (2) The Antagonist. Book Two is called Prescription and includes four chapters, each called The Revolutionist's Handbook. The so-called

diagnostic chapters contain more startling and surprising subject matter and interpretations. The chapters on prescription are mild by comparison.

The book, advises the author, represents a feeble attempt on his part to write a prescription for living for use by all. No one reading the book will consider it a feeble attempt. The book is written in strong language and it is more than outspoken—it is often condemnatory of many of the things considered to be proper now in the field, and is written with seeming aim to disturb and upset people into action and get them tense enough to express interest in doing something about it. A cult of passivity and surrender is exactly what he is fighting against and he fights openly and strongly. For the first many chapters, the book sounds more like that written by a prophet of doom or the chronicler of psychological disaster than one who is making a feeble attempt. The very first sentence of the prologue is illustrative. It reads "the plight of modern man is pitiful . . . Now, with this knowledge in his grasp, he finds it being misunderstood or conscientiously misapplied. If ever there was cosmic tragedy, this is it." This error of prescribing for adjustment, in his words, amounts to a literal perversion of a most important, if not the most important, area of knowledge to an instrument for the continued debasement, exploitation and enslavement of man.

Antagonists with good professional status, he has many. He particularly picks on Franz Alexander for what he considers to be his emphasis in therapy, namely, domestication; Karen Horney for representing only the upholding of feminine psychology; Harry Stack Sullivan for seeming to be too satisfied with happiness based on conformity and, in that sense, being an advocate of reaction; Carl Rogers, whose non-directive psychotherapy he refers to as a "potpourri of ventilation, reassurance, insight and will therapy." He is even more vehemently opposed, and spends more time considering with hostility, what he refers to as the 'shockiatrists'. He is also critical of psychosurgery and its indiscriminating use, particularly. But he does have some protagonists. He reacts favorably to Abraham Myerson, Bertrand Russell, Norbert Weiner, Roland S. McCullough. And, after damning many psychoanalysts, he comes out with the conclusion that the most effective of all psychotherapeutic tools for accomplishment of the aim of treatment is psychoanalysis. But he adds that the

analyst has to assume the responsibility for the philosophy with which the patient leaves his analysis, and that the philosophy should be that of rebellion on which the fundamental premise derives from the last three attributes of the rebellious person: responsibility, employment and tension.

That adjustment is not necessarily a criterion of a good life, a wholesome personality or integrity, is not exactly a new concept in the field of mental hygiene. William Healy, psychiatrist, Clifford Shaw, sociologist, and other students of delinquency have differentiated between integration as against adjustment, and have written at length that adjustment to the exploited surroundings of a delinquency area does not represent good mental health, but merely means conforming to delinquency. It probably is a fact that all too often in the more academic texts, adjustment has been described in a manner which justified criticism, but whether this justifies this kind of alarming consideration is a question mark. The end in view is to some extent justified; the means are extremely questionable. But, regardless, this is an interesting book to read, which emphasizes the attitude that is deserving of some attention, although it is expressed in over-generalized, exaggerated, and over-revolutionary terms. As such, it can make for valuable reading if taken as an expression of protest rather than as a statement of facts.

H. MELTZER

Psychological Service Center
St. Louis, Missouri.

CARL A. WHITAKER AND THOMAS P. MALONE. *The Roots of Psychotherapy*. New York: The Blakiston Company, 1953, pp. 236.

In *The Roots of Psychotherapy* two professors from Emory University's department of psychiatry, one a Ph.D. and one an M.D., make an attempt to develop a frame of reference to serve as a basis for considering features of the treatment process which are common to all forms of psychotherapy regardless of differences in techniques or in original orientation. This they do on the assumption that in recent years psychotherapy has become more and more a separate discipline dealing with a recognizable process. The content of the book is presented in three large sections: Foundation, Process, and Techniques. The section on Foundation considers the biological

basis of psychotherapy, implicit and explicit psychotherapy, and dynamic psychiatry and psychotherapy. The section on Process includes chapters on the patient as a person, the process of psychotherapy, anxiety and psychotherapy, the therapist as a person, patient-vectors in the therapist, and restatement of the problem of psychotherapy. In the section on Techniques are presented some of the techniques found useful by the authors. The book is intended to be for anybody, no matter what his training, who has the experience of being a patient or who has been a patient himself. The impression the authors express in the preface is that such people should have little difficulty comprehending the essential meaning of the book. The book contains a glossary of terms as well as a bibliography, but does not contain an index.

Of interest to clinicians will be the chapter on some techniques in brief psychotherapy. Here will be found a brief consideration of the various stages of therapy—beginning phase, symbolic phase, and ending phase. Here too they will find a consideration of techniques of administration in therapy. Technique is defined by them as “an interpersonal operation deliberately used by the therapist, the function of which is to transpose social, latent affect in both participants into deeper, manifest affect in order to catalyze the affective and symbolic process of psychotherapy.” This may be a good definition, but not easily understood by a layreader. Presented here also are techniques for expanding the fantasy component of therapy, this on the contention that adult oriented fantasy frameworks are less significant as methods of communication than are the symbolic forms which the culture provides. Also considered in brief are pitfalls of psychotherapy.

The authors show a ready familiarity with not only psychiatric and psychological literature but also related philosophical literature. Illustrative references include the writings of Rank, Ruesch, Schilder, H. S. Sullivan, Kubie; Rogers, Lewin, and Piaget; and Einstein, Morris Cohen, Bridgeman, and Whitehead. Sane and professionally sound the book is. An honest attempt to build up and develop a frame of reference for consideration of therapy for practice and in research purposes it also is. And valuable to professional therapists it can be. For popular purposes the writing just isn't clear enough, particularly in the early chapters.

H. MELTZER

Psychological Service Center, St. Louis, Mo.

FRANCES LLOYD. *Educating the Sub-normal Child*. New York: Philosophical Library, 1953, pp. 148. \$3.75.

JOHN C. TOWNSEND. *Introduction to Experimental Method*. New York: McGraw-Hill Book Company, 1953, pp. 220. \$4.00.

Here are two books interested on very different levels in the educational methods outlined and recommended for the two kinds of problems: the one, for backward and subnormal children; the other, in the teaching of college students in the new method of learning known to the human mind; namely, learning by experiment. Both books have significant agreement in one of the fundamentals of teaching, that of getting the learner to participate actively in the learning.

Lloyd reports work that she has been doing in a special school which the Foreword maintains is "an ordinary school and not a show school of the type to which authorities like to direct visitors." (p. v).

The information and discussions include the problem of the criterion by which E. S. N. (educationally sub-normal) children should be selected for special school treatment; social background and attitudes of parents; social training in the special school; misfits in the special E. S. N. school; activities in the special E. S. N. school; what success means in this kind of school; and three case studies.

The value of the book lies not in the fact that there is anything particularly new reported; but rather in the brief story of what appears to be excellent work in an English school on this difficult level of education. It may also add to the encouragement of those who are now in many places considering the organization of such schools or classes, and who desire to know the judgment of one who has had long experience in the work.

In the midst of what seems to be useful information and good educational judgment, there are amazing statements about IQ's. The IQ of one of the cases was reported as 42 at the age of seven, and 89 at the age of twelve. (p. 81). And a 'doctor' was reported (p. 82) as suggesting that under proper conditions, the IQ might rise another twenty points! One wonders what psychological examination was used; what was the competence of the examiner; and particularly, was the child in proper condition when he was examined so that it was possible to obtain an accurate measure

and intelligence quotient? Especially, we should like to know if the author means to say that the methods of the school actually raised an imbecile so near to the intelligence of a normal child. If this can be done, educators will want to know a great deal more about the methods used in this school.

A useful index completes the book.

Quite as important, and for very different reasons, is the volume of Townsend. Its contribution is not to teaching in one or a few of the sciences, but to teaching in all fields of science. Its purpose is to demonstrate how to get away from 'cook-book' methods, away from mere following of instructions, to insight and actual participation in the performing of scientific experiments.

"The emphasis in this book is placed on preparing the student to think along lines of the development of sound research designs rather than the successful execution of standardized traditional experiments. In this way, it is believed that the beginning student may be exposed to the elements of theory and method necessary for the understanding of the performance of experiments. Equally important, the student will be given opportunity to become sensitive to the ways of critical thinking early in his contact with science and not be forced to labor in darkness and boredom until he becomes a graduate student." (p. vii).

The main parts of the book include theory of experimentation; design and conduct of experiments; interpretations and conclusions; and applications of the experimental method.

In such a book there is the very difficult problem of selection of material, especially, since it is a relatively new kind of book. In its ensemble it is new. In detail it considers scientific versus non-scientific procedure; the locating and simplifying of problems; hypotheses and variables; control of the experiment and use of apparatus; the conducting of the experiment; the use of logic and of statistics; the making of the report including the making of tables, but not enough about the making of tables; pictures and brief descriptions of apparatus; and the appendix includes four of the most commonly used statistical tables.

It is easy to point out details which one would suggest for inclusion in the volume. The reviewer is inclined to think that the book would be more interesting and useful to the student if there were more concrete psychological subject matter to balance the amount of abstract and methodological material. From one point of

view the book is more useful to the instructor than to the student because the instructor already has the concrete material on the basis of which the logic and methodology can be thought out. The instructor can supply the concrete material and may do so by using a supplementary text. Except for some mention of apparatus there is practically nothing in the book relating to animal psychology.

It is believed that the book is genuinely useful for the purpose of helping the student to learn science in place of the continued demand to follow instructions which come from the nursery school up. It is also believed that the use of such methods as suggested by Townsend could vastly improve the teaching of science in all departments of colleges and universities; and a somewhat simple application could be made to improve science teaching in the high schools. But it will need more concrete subject matter material.

The book concludes with a useful index, pp. 215-220.

A. S. EDWARDS

The University of Georgia

PUBLICATIONS RECEIVED

- ANNE ANASTASI. *Psychological Testing*. New York: The Macmillan Co., 1954, pp. 682. \$6.75.
- GLADYS ANDREWS. *Creative Rhythmic Movement for Children*. New York: Prentice-Hall, Inc., 1954, pp. 198. \$4.75.
- JOSEPH BARRELL. *A Philosophical Study of the Human Mind*. New York: Philosophical Library, 1954, pp. 575. \$6.00.
- FRANZISKA B. GORTEN. *Die Psychologie der Menschenbehandlung*. Third Edition. Zurich, Switzerland: Rascher & Cie, pp. 40. \$4.75.
- SAMUEL J. BECK. *The Six Schizophrenias: Reaction Patterns in Children and Adults*. New York: The American Orthopsychiatric Association, Inc., 1954, pp. 238. \$5.00.
- RALPH F. BIRDIE. *After High School What?* Minneapolis: University of Minnesota Press, 1954, pp. 240. \$4.25.
- TRYGVE BRAATØY. *Fundamentals of Psychoanalytic Technique*. New York: John Wiley & Sons, Inc., 1954, pp. 404. \$6.00.
- DOUGLAS W. BRAY. *Issues in the Study of Talent*. New York: King's Crown Press, 1954, pp. 65. \$2.00.
- EVA BURMEISTER. *Roofs for the Family: Building a Center for the Care of Children*. New York: Columbia University Press, 1954, pp. 203. \$3.25.
- SIR CYRIL BURT. *The Causes and Treatment of Backwardness*. New York: Philosophical Library, 1954, pp. 128. \$3.75.
- TIMOTHY CANNON. *Psychology: The Unity of Human Behavior*. Boston: Ginn & Co., 1954, pp. 482. \$4.75.
- LEONARD CARMICHAEL, Editor. *Manual of Child Psychology*. Second Edition. New York: John Wiley & Sons, Inc., 1954, pp. 1295. \$12.00.
- ELY CHINØY. *Sociological Perspective: Basic Concepts and Their Application*. Garden City, N. Y.: Doubleday & Co., Inc., 1954, pp. 58. (paper) \$0.85.
- LEE J. CRONBACH. *Educational Psychology*. New York: Harcourt, Brace & Co., 1954, pp. 628. \$7.50.
- Decision-making and American Values in School Administration*. New York: Bureau of Publications, Teachers College, Columbia University, 1954, pp. 90. (paper) \$2.00.
- RUTH E. ECKERT AND ROBERT J. KELLER, Editors. *A University Looks at Its Program*. The Report of the University of Min-

nesota Bureau of Institutional Research, 1942-1952. Minneapolis: The University of Minnesota Press, 1954, pp. 223. \$4.00.

HAROLD KENNETH FINK. *Mind and Performance: A Comparative Study of Learning in Mammals, Birds, and Reptiles*. New York: Vantage Press, Inc., 1954, pp. 113. \$3.00.

DOUGLAS H. FRYER, EDWIN H. HENRY, AND CHARLES P. SPARK. *General Psychology: Survey of Basic Principles*. New York: Barnes & Noble, Inc., 1954, pp. 300. (paper) \$2.50.

MARVIN D. GLOCK. *Psychology of Learning*. Boston: Houghton Mifflin Co., 1954, pp. 307. (paper) \$2.50.

CARTER V. GIBBS AND DOUGLAS E. SCATES. *Methods of Research*. Appleton-Century-Crofts, Inc., 1954, pp. 920. \$6.00.

KENNETH W. HODGSON. *The Deaf and Their Problems*. New York: Philosophical Library, 1954, pp. 364. \$6.00.

J. ANTHONY HUMPHREYS AND ARTHUR E. TRAXLER. *Guidance Services*. Chicago: Science Research Associates, Inc., 1954, pp. 438.

ARTHUR T. JERSILD. *Child Psychology*. Fourth Edition. New York: Prentice-Hall, Inc., 1954, pp. 676. \$6.00.

GEORGE R. KLARE AND BYRON BUCK. *Know Your Reader: The Scientific Approach to Readability*. New York: Hermitage House, Inc., 1954, pp. 191.

CLAUDE KOHLER. *Les Deficiences Intellectuelles, Chez l'Enfant*. Paris: Presses Universitaires de France, 1954, pp. 222. (paper) \$2.50.

HENRY CLAY LINDGREN. *Mental Health in Education*. New York: Henry Holt & Co., 1954, pp. 561. \$4.75.

NORMAN ROBERTS LYKES. *A Psychological Approach to Accidents*. New York: Vantage Press, Inc., 1954, pp. 136. \$2.95.

ROGER MARIER. *Social Welfare Work in Jamaica*. Paris (19 Avenue Kleber): UNESCO, 1953, pp. 166. (paper) \$2.50.

WILLIAM E. MARTIN AND CELIA BURNS STENDLER. *Readings in Child Development*. New York: Harcourt, Brace & Co., 1954, pp. 513.

V. J. MCGILL. *Emotions and Reason*. Springfield, Ill.: Charles C. Thomas, 1954, pp. 122. \$3.25.

- G. MIALARET. *Nouvelle Pedagogie Scientifique*. Paris: Presses Universitaires de France, 1954, pp. 123. (paper)
- DORIS M. ODLUM. *Psychology, the Nurse, and the Patient*. Revised Edition. New York: Philosophical Library, 1954, pp. 168. \$7.50.
- HAROLD B. PEPINSKY AND PAULINE NICHOLS PEPINSKY. *Counseling: Theory and Practice*. New York: The Ronald Co., 1954, pp. 307. \$4.50.
- HAROLD R. RYDEN AND CLELLEN L. MORGAN. *Introduction to Educational Psychology*. New York: Harper & Brothers, 1954, pp. 435. \$4.50.
- ANDRE LA RIVIERE. *La Nevrose Maladie Tropicale et Somnolence*. I. pp. 266. *La Nevrose Cette Grande Misere Humaine*. II. pp. 266. *La Nevrose Rempart de la Maladie*. iii. pp. 260. Montreal (3426 Ave. Marcil, N. D. G.): Les Editions Psychologiques Engr., 1953. (paper)
- C. C. ROSS AND JULIAN C. STANLEY. *Measurement in Today's Schools*. New York: Prentice-Hall, Inc., 1954, pp. 485. \$5.00.
- DAGOBERT D. RUNES. *Letters to My Daughter*. New York: Philosophical Library, 1954, pp. 131. \$2.50.
- DAVID H. RUSSELL. *The Dimensions of Children's Meaning Vocabularies in Grades Four through Twelve*. Berkeley, Calif.: University of California Press, 1954, pp. 100. (paper) \$1.25.
- RICHARD L. SCHANCK. *The Permanent Revolution in Science*. New York: Philosophical Library, 1954, pp. 112. \$3.00.
- HENRY P. SMITH. *Psychology in Teaching*. New York: Prentice-Hall, Inc., 1954, pp. 466. \$4.95.
- GEORGE D. STRAYER. *A Design for the Administration of Public Education with Particular Applications to California*. Stanford, Calif.: Stanford University Press, 1954, pp. 31. (paper)
- JANE WARTERS. *Techniques of Counseling*. New York: McGraw-Hill Book Co., 1954, pp. 384. \$4.75.
- JOHN P. ZUBEK AND PATRICIA ANNE SOLBERG. *Human Development*. New York: McGraw-Hill Book Co., 1954, pp. 476. \$6.00.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 45

December, 1954

Number 8

A STUDY OF TRANSFER: SEX DIFFERENCES IN THE REASONING PROCESS

MAX MARTIN KOSTER

College at F

A study was undertaken to determine whether sex differences in the transfer of training could be discovered through a statistically controlled experiment. Transfer for this study was defined as the successful application of skill or knowledge in adjusting to a new situation. The amount of transfer to the new situation was measured by the difference between the actual performance and the expected performance. The actual performance was the attainment after skill or knowledge to be transferred had been learned. The expected performance was what the attainment would have been had the skill or knowledge to be transferred not been learned.

For the purpose of delimiting the scope of the investigation, transfer of training was restricted to the successful application of generalizations only. Sex differences in the amount of transfer were measured by comparing scores obtained in the new situation with scores predicted on the basis of knowledge acquired previous to training.

The technique used in this investigation differed from the currently popular techniques for measuring sex differences in two respects: first, this technique used a statistical control on previous knowledge; second, while the prevalent techniques for measuring sex differences used neuter items in preference to biased items, the present investigator used biased items in preference to neuter items.

Because intelligence and transfer of training are intimately related, the present writer thought it advisable to examine experiments on sex differences in intelligence. The results of this examination revealed the general belief that there is no sex difference in intelligence. However, the writer feels that this belief is based

upon sound evidence. The prevalent intelligence tests are invalid when used for the measurement of sex differences in intelligence because of the nature of the items. The accepted procedure for selecting items is by the method of 'fairness to boys and girls'. This method results in items, and in turn tests, that do not and cannot disclose sex differences in total scores. Surely such neuter items are not appropriate for detecting sex differences.

As sex differences in intelligence have not been accurately measured, any valid information as to sex differences in the transfer of training is wanting. This is the basic problem of measuring sex differences in mental ability. pp. 435. \$4.00.

The major problem, 'Do girls transfer more than boys?' was broken down into two problems, 'Do girls transfer more than boys?' and 'Do boys transfer more than girls?' Each problem requires a special experimental design with different source materials. As a preparation for the selection of the source materials, numerous surveys on reading habits and reading preferences were consulted. These surveys indicated that boys prefer to read and actually do read more about Science; whereas girls prefer to read and actually do read more Home Economics. Accordingly, Home Economics material was used as source material for investigating whether boys transfer more than girls. Should boys be found to obtain higher scores on Home Economics, the higher scores could not be explained away on the basis of sex differences in previous knowledge.

The writer investigated transfer of training under two different conditions when hints are given and when hints are not given. A total of three hundred boys and three hundred girls drawn from Quincy, Massachusetts; Boise, Idaho; and Charleston, South Carolina were used.

The design for investigating whether boys transfer more than girls when no hints are given contained three types of items: (1) deduction items, (2) information items, (3) items on principles. The deduction items were used as a measure of transfer of training. The information items were used as a check on whether the deduction items were biased in favor of one sex. The items on the principles were used as a measure of previous knowledge.

The items were selected in accordance with the following specifications:

- 1) Items selected for the first section of the test were based on

Science material. Items selected for the second section were based on Home Economics material.

2) The Home Economics items were presented to several females for approval as to whether or not the items were as fair to girls as to boys. Fairness in this instance was based upon interest, utility and preoccupation rather than upon obtained scores.

3) Items were chosen which seemed to be new, interesting and challenging and to have application in the pupils' daily lives.

4) The deduction items that were selected and be answered by the application of the principles described in these items. These principles could be used inductively from the helpful suggestions.

5) All the items were written in clear, concise language that could be understood by pupils several grades below the eleventh. Thus, clarity was assured for the test group which was composed of pupils selected from the eleventh grade.

SAMPLES OF ITEMS FROM SCIENCE MATERIAL

General Information Item	
Heat tends to make air expand.....	T F
Item on Principle	
Still air is a good conductor of heat.....	T F
Deduction Items	
On a cold day, a loose-fitting glove will keep your hand warmer than a tight-fitting one.....	T F
Insulating material is made loose and fluffy so it will be easier to handle.....	T F
Helpful Information that was given between pretest and post-test	
Less heat is lost through hollow building blocks than through solid ones because "still air" does not conduct heat as well as cement does.	

SAMPLES OF ITEMS FROM HOME ECONOMICS MATERIAL

General Information Item	
As a rule, bluing is used to make clothes appear blue.....	T F
Items on Principle	
A fiber of cloth is weakened by twisting and bending.....	T F
Deduction Items	
For better care, place fresh linen upon the linen already in the drawer.....	T F
Washing clothes will wear them out.....	T F
Helpful Information given between pretest and post-test	
The wearing qualities of textiles is sometimes tested by counting the number of times the fiber of the cloth can be bent and straightened.	

The test was administered to over twelve hundred eleventh-grade pupils. From this group, six hundred test papers were selected for the final statistical analysis. These papers were selected on the basis of age and type of high-school program. Test papers of pupils between the ages of fifteen and seventeen, and a proportionate number of test papers belonging to college and non-college preparatory pupils were used.

The test papers selected from Quincy, Massachusetts, were screened further, leaving off the boys' and girls' papers only when the tenth member of the pair had obtained the same reading comprehension score, plus any *deduction items* in the scores on the deduction items, if found, could not be attributed to sex differences in reading comprehension.

However, reading comprehension is only one of many components that might account for sex differences in the scores on deduction items. In order to be reasonably certain that sex differences in scores on deduction items indicated sex differences in transfer of training, other possible explanations for sex differences in the scores must be eliminated or at least minimized. Many possible explanations for sex differences may be eliminated by determining whether or not the boys and girls used in this study were unique in a way that would invalidate the general conclusion that boys transfer more than girls.

How unique is the sample group of boys and girls? This question was investigated by comparing boys and girls on the basis of standardized tests, of teachers' marks on school subjects and teachers' ratings of pupils' personality.

The pupils in Quincy took the Pintner Intermediate Intelligence Test when in the tenth grade, and the mean IQ score for the girls (112.5) was found to be slightly higher than the mean score for the boys (110.0). Hence, any superiority of boys scores on deduction items could not reasonably be attributed to superiority in intelligence.

On examining school marks of the experimental group, the girls were found to obtain slightly higher marks than the boys in both Mathematics and Science. However, teachers' marks on school subjects are known to show a 'halo effect' in favor of girls. This 'halo effect' may be partly due to sex differences in personality.

On examining the personality ratings, the girls were found to receive a slightly higher rating on 'initiative' and a significantly

higher rating on 'applicability'. Thus, if boys received higher scores on deduction items, the higher scores could not be explained away on the basis of personality traits such as initiative or applicability.

Only on the Arithmetic Reasoning test did boys obtain higher mean scores than girls and, even then, the difference between the scores was not significant. It may well be that Arithmetic Reasoning has elements of deduction and transfer.

The sample of pupils selected for this investigation was not found to be unique in a way that would invalidate the conclusion that boys transfer more than girls.

In adapting a study to this investigation, the major question, "Who transfer more, boys or girls?", was broken down to "Who transfer more when helpful suggestions are given, boys or girls?" and "Who transfer more when helpful suggestions are not given, boys or girls?" This breakdown makes it possible to isolate the effect suggestion may have on test scores. To account for sex differences in reading preferences, reading habits, life goals and previous knowledge, the two questions were further broken down into four:

- 1) Do girls transfer more Science material than boys when helpful suggestions are not given?
- 2) Do boys transfer more Home Economics material than girls when helpful suggestions are not given?
- 3) Do girls transfer more Science material than boys when helpful suggestions are given?
- 4) Do boys transfer more Home Economics material than girls when helpful suggestions are given?

Table 1 is the design for investigating the question, "Do girls transfer more Science material than boys when helpful suggestions are not given?" The schema is paralleled by the actual statistical results.

Do girls transfer more Science material than boys when helpful suggestions are not given? The answer seems to be 'No'; at least, the requirements of the proof were not fulfilled. The first requirement that girls should not obtain higher scores on Science information was fulfilled. The data do not support the view that boys have less knowledge of the specific type of Science material used in the test.

The second requirement that girls do not obtain higher scores on

TABLE 1.—DO GIRLS TRANSFER MORE THAN BOYS?

Tests on	Requirements for proof	Actual results
1) Information items (Science material)	Girls should not obtain higher scores	Girls did not obtain higher scores $P =$.00,00+
2) Items on principles (Science material)	Girls should not obtain higher scores	Girls did not obtain higher scores $P =$.0035
3) Deduction items (Science material)	Girls should obtain higher scores	Girls did not obtain higher scores $P =$.00,000

principles pertinent to deduction items was also fulfilled. The findings support the view that the boys' scores on science principles were significantly higher than the girls' scores. Therefore, if girls did obtain higher scores on deduction items than boys, the higher scores could not reasonably be attributed to sex differences in previous knowledge.

The third and crucial requirement that girls obtain higher scores on deduction items on Science was not fulfilled. The data indicate that the boys obtained significantly higher scores on Science items ($P = .00,000$) and, therefore, these results can not support the view that girls transfer more than boys.

"Do boys transfer more Home Economics material than girls when helpful suggestions are not given?" may be investigated with a new design. Table 2 is the schema with the actual statistical findings.

Do boys transfer more Home Economics material than girls when helpful suggestions are not given? The answer is 'Yes'. All three requirements were fulfilled. The data support the view that girls have more knowledge of the specific type of material used in the test. Hence, any superiority boys may have over girls on deduction items could not reasonably be attributed to a superiority in previous knowledge of Home Economics.

The second requirement that boys do not obtain higher scores on principles pertinent to deduction items was also fulfilled. The data support this view. Therefore, any superiority of boys' scores over girls' scores on deduction items could not reasonably be

TABLE 2.—DO BOYS TRANSFER MORE THAN GIRLS?

Tests on	Requirements for proof	Actual results
1) Information items (Home Economics material)	Boys should not obtain higher scores	Boys did not obtain higher scores $P = .00,000$
2) Items on principles (Home Economics material)	Boys should not obtain higher scores	Boys did not obtain higher scores. Girls go slightly higher $P = 30$
3) Deduction items (Home Economics material)	Boys should obtain higher scores	Boys did obtain higher scores. $P = .00,000$

attributed to a superiority in knowledge of principles useful in figuring out the deduction items.

The third and crucial requirement was that boys obtain significantly higher scores on deduction items drawn from Home Economics material when helpful suggestions are not given. This requirement was fulfilled at a level of significance of $P = .00,000$.

All three requirements for proving that boys transfer more than girls were fulfilled. The findings, therefore, support the statement that boys transfer more than girls when helpful suggestions are not given.

"When helpful suggestions are given, do girls transfer more in Science material?" may be investigated by the design in Table 3.

TABLE 3.—GIVEN HELPFUL SUGGESTIONS, DO GIRLS TRANSFER MORE THAN BOYS?

Tests on	Requirements	Results
1) Information items (Science material) Pre-test followed by post-test	Girls should not gain more than boys	Girls did gain slightly more than boys
2) Deduction items (Science material) Pre-test followed by helpful suggestions and post-test	Girls should obtain higher adjusted post-test scores	Girls did not obtain higher adjusted post-test scores

Do girls transfer more Science material when helpful suggestions are given? The answer is 'No'. Neither of the requirements was fulfilled. The purpose of the first requirement was to determine whether or not girls gain more than boys by the 'practice effect'. The practice effect was measured by the difference between the pre-test and the post-test scores on the information items. In this experiment, the girls were found to gain slightly more than the boys by the 'practice effect' for both Science material and Home Economics material. The girls' gain for Science material was 1.35 points while the boys' gain was 0.63 points. For Home Economics material the girls' gain was 1.66 points while the boys' gain was 1.23 points. The first requirement was not satisfactorily met.

Neither was the second and crucial requirement satisfactorily met. After helpful suggestions were given and after allowances were made for the effect of previous knowledge of the deduction items, the boys were found to obtain significantly higher scores on deduction items on Science $P = .0014$. 'Allowances' were made by discounting the effect of pre-test score upon post-test score through the analysis of covariance.

"When helpful suggestions are given, do boys transfer more in Home Economics material?" may be investigated by the design in Table 4.

Do boys transfer more in Home Economics material when helpful suggestions are given? The answer is 'Yes'. Both requirements were fulfilled.

TABLE 4.—GIVEN HELPFUL SUGGESTIONS, DO BOYS TRANSFER MORE THAN GIRLS?

Tests on	Requirements	Results
1) Information items (Home Economics) Pre-test followed by post-test	Boys should not gain more than girls	Boys did not gain more than girls
2) Deduction items (Home Economics) Pre-test followed by helpful suggestions and post-test	Boys should obtain higher adjusted post-test scores	Boys did obtain higher adjusted post-test scores $P = .00,000$

Girls were found to gain slightly more than boys by the 'practice effect'. Evidence for this statement has been given in the discussion of the third question, and therefore the first requirement of the proof has been fulfilled.

Evidence with a $P = .00,00$ support the view that boys obtain higher adjusted scores on the deduction items on Home Economics material when helpful suggestions are given.

Hence, with both requirements fulfilled, it may be stated that boys transfer more than girls when helpful suggestions are given. By synthesizing the above conclusion that boys transfer more than girls when helpful suggestions are given with the previously stated conclusion that boys transfer more than girls when helpful suggestions are not given, one is led to the general conclusion that boys generally transfer more than girls.

The conclusion that boys generally transfer more than girls is based on the fact that boys obtained significantly higher scores on deduction items in Home Economics material before suggestions were given, and also after helpful suggestions were given.

The fact that boys obtained significantly higher scores on deduction items in Home Economics could not be attributed to a superiority in intelligence, because the boys did not receive higher intelligence test scores.

The higher scores of the boys could not be due to the fact that there was a superiority in 'applicability' and 'initiative', as the girls had received significantly higher ratings from their teachers on the personality trait labeled 'applicability' and 'initiative'.

The higher scores that boys received could not be explained on the basis of reading preference and reading activities because surveys have shown that girls prefer to read and actually do read more about Home Economics than boys.

The boys' higher scores on deduction items could not be explained on the basis of previous knowledge because the scores of the boys were lower than those of the girls on information drawn from Home Economics material.

The difference in the scores could not be explained on the basis of knowledge of the principles pertinent to the deduction items. The girls' mean score on the principles was slightly higher than the boys' mean score.

The boys' greater gain in scores from pre-test to post-test on

deduction items cannot be attributed to superiority in the practice effect. Girls were found to gain slightly more than boys by the practice effect.

A superiority in the scores could not be attributed to a sex difference in reading comprehension because the boys and girls used in the investigation were experimentally matched on the basis of reading comprehension.

In summary, it may be said that since the boys' superior mean score on the deduction items is not due to intelligence, to previous knowledge, to reading comprehension, to reading preferences or reading activities, to personality traits such as applicability and initiative, to practice effect, or knowledge of pertinent principles, it is reasonable to conclude that the higher scores on the deduction items are due to boys' superior ability to transfer.

PERCEPTION IN NUMBER SKILLS—A STUDY IN TACHISTOSCOPIC TRAINING¹

JOHN L. PHILLIPS, JR.

Department of Educational Psychology, University of Utah

A perusal of the literature on tachistoscopic training discloses what appears to the writer as a single basic issue, with two major problems growing out of it. The basic issue is between two classes of theory, called in the psychology of reading, 'peripheral' and 'central'. (These terms can be extended to refer to any skill involving the use of the higher mental processes). The two problems growing out of it are (1) can peripheral training improve performance in these skills? and (2) can central training improve performance in these skills?

The answer to the latter question is so obviously affirmative as to preclude any extended discussion. No one would question the necessity of concept formation as a prerequisite to reading, for example, or to recognition of aircraft, or to picking out the imperfect items on an assembly line. The other proposition is not so apodictic; it is the question that has provided stimulus for the present study. Can peripheral training improve performance in mental skills?—more specifically, can this type of training use short-exposure methods to advantage?

It can be stated with considerable assurance that performance on perceptual tests of various kinds can be improved by practice in the reproduction of briefly-exposed stimuli. But does this improvement in laboratory tasks transfer to skills involved in scholastic and other more utilitarian performances?

Again the answer is seemingly affirmative; but when only those studies are considered that include trained control groups—i.e., groups trained by one or more alternative methods—it becomes apparent that relatively little evidence is available.

Dallenbach, working with normal (2) and later with feeble-

¹ Adapted from a thesis submitted to the Graduate School of the University of Utah and to the Department of Educational Psychology in partial fulfillment of the requirements for the Ph. D. degree. The study was supported by a research grant from the University Research Committee and was completed under the chairmanship of Professor R. DeVerl Willey.

minded (3) school children found a general increase in scholastic achievement as compared with a traditionally trained group; Renshaw (8, 10), and Melcer and Brown (6) confirmed him. Renshaw (9) also reported a rather startling superiority for tachistoscopic training in reading at the college level. Brown (1) used the instrument to advantage in teaching spelling to seventh-graders. Winger's (13) college students in beginning typing and the Navy's (Neville, 7) aircraft recognition trainees performed better when trained by tachistoscopic methods.

On the other hand, Weber (13), Sutherland (12), and Freeburne (4) set up training programs in reading for college students in such a way that the gains of the tachistoscopically-trained groups could be compared with those of groups trained by alternative methods, and although they found large gains being scored by their subjects, these gains were shared equally by all groups. Hoyt, *et al.* (5) tried the same thing in a college art course, and with the same result.

Taken as a whole, then, the evidence indicates a considerable value for short-exposure methods; but the evidence is far from conclusive, for there are important experiments on record in which the results are definitely negative. These methods have failed to establish decisively their superiority over the more traditional techniques with which they must in the last analysis compete for preëminence as clinical and classroom procedures.

AN EXPERIMENT IN ARITHMETIC TRAINING

Computational skill, defined as the ability to respond consistently and correctly to the presentation of certain number combinations, was selected to be the focus of the investigation reported in this paper. An experiment was conducted in the fourth grade of a Salt Lake City, Utah, public school (first experiment) and then repeated in another school in another part of the city (second experiment). The primary purpose was the testing of two hypotheses: (1) that tachistoscopic training is effective in the teaching of computational skills, and (2) that tachistoscopic training is more effective than a more conventional, paper-and-pencil approach. (Other problems were considered along the way, and will be discussed in the section on conclusions, but they were not the central objectives.)

Three experimental groups and a control group were equated in terms of intelligence (mean SI^2 , approximately one hundred), grade in school (fourth), sex, visual acuity, and number of training days present. Inasmuch as the classes had already been organized when the experiment began, it was necessary to do this equating of groups by casting out cases after the experiment had been completed. N for each of the equated groups in the first experiment was ten (five girls and five boys); while in the second experiment there were five girls and nine boys, or fourteen in each group. In the first experiment Group IV, the control group (see below, Group IV), was eliminated as a result of a procedural error by the teacher in charge; in the second, all four groups were studied. In both experiments, the first half of the training was supervised by the regular classroom teacher (the same teacher for all groups), while in the latter half, the experimenter was in charge. Conditions of training were as follows:

Group I: a paper-and-pencil procedure in which the practice sheets were like the pages in a workbook.

Group II: an adaptation of the tachistoscopic method developed by Willey and Billington and published in the *Keystone teaching manual*. It consisted primarily of reproducing (on paper) complete number combinations (both sides of the equation sign) when these were flashed onto a screen.

Group III: developed by the writer for these experiments; it also made use of the tachistoscope, but in a different way. Instead of flashing the entire combination onto the screen in a single exposure, the teacher flashed first the problem and then, after a short pause, the answer. In the 3-sec. interim between 'problem flash' and 'answer flash', the subject attempted to recall the correct answer and record it on paper.

Group IV: planned as a control in which no special training would be given on the set of combinations specified for training of the other groups. This control group merely continued with its regular classroom work, encountering the experimental combinations only as they happened to occur in textbook assignments and with no special attention being directed toward them.

The amount of time spent in arithmetic classes per day was the

² The SI (sigma index) is a standard score in a distribution of Salt Lake City children's scores with a mean of 100 and a standard deviation of 20.

same for all groups. The amount of time spent per period in working on the experimental combinations³ was equal for all excepting the control group, where it was not measurable but was known to be very small.

The training series comprised twenty multiplication and twenty division combinations selected primarily from the course of study for the next higher grade (fifth). These were typed onto transparent slides and projected in .11 sec. exposures⁴ by a Keystone tachistoscope—an overhead projector with a diaphragm shutter attachment.

Assessment of arithmetic gains made during training was done by means of a battery of arithmetic tests given before and after training, which had been constructed especially for the purpose and which presented the trained combinations in each of four ways: (1) horizontal, original order—exactly as presented in training; (2) vertical, original order; (3) horizontal, reversed order; and (4) vertical, reversed order.⁵ The effect of training on perceptual ability was tested by asking subjects also before and after training to reproduce on paper numbers of varying length (one through nine digits) when flashed at .03 sec. Slides were made in the same way as were those in the training series, and exposures were made by the same equipment.

The data were analyzed into five categories: Perception;⁶ Horizontal, Original Order;⁷ Multiplication;⁸ Division;⁹ and Arithmetic, Total¹⁰; these were converted into difference scores ('after training' minus 'before training') and subjected to analysis of variance in all categories.

³ The length of practice periods varied, but they averaged approximately ten minutes for the two experiments. In the first experiment there were fourteen daily practice sessions; in the second there were twelve.

⁴ Calibrated "1/25 sec." on the face of the Flashmeter.

⁵ For example: (1) $9 \times 7 =$; (2) $\frac{9}{63}$; (3) $7 \times 9 =$; (4) $\frac{7}{63}$; or (1) $63 \div 9 =$; (2) $9/63$; (3) $63 \div 7 =$; (4) $7/63$

⁶ As measured by the test described in the paragraph above.

⁷ Including both multiplication and division.

⁸ Including both horizontal and vertical, both original and reversed.

⁹ Including both horizontal and vertical, both original and reversed.

¹⁰ Including multiplication and division, horizontal and vertical, and original and reversed.

DIVERGENT RESULTS FROM THE TWO EXPERIMENTS

Results from the two experiments can be characterized as convergent (in basic agreement) or divergent (in disagreement). The latter are important insofar as they throw doubt on the former.

Before the data were converted into difference scores, three of the pre-training measures were singled out for examination of their interrelations; these three (the total score from all the arithmetic tests, the perception test score, and the intelligence test score) were selected because of the theoretical significance that attaches to relations among them.

On the basis of the theory advanced by proponents of tachistoscopic training, there should be some correlation between performance in any intellectual function and the perceptual skill which, according to the peripheral theory, underlies it. For example, arithmetic test scores and perception test scores should be positively correlated, as should intelligence test scores and arithmetic test scores. Data from the two experiments do not agree on this point: the first experiment showed no correlation, the second produced coefficients of $r = .30$ and $r = .25$, respectively, for the two correlations. Both are beyond the .05 level of confidence. Since the total N in the first experiment was only thirty, while that for the second was fifty-six, differences in significance levels may be due in part to the difference in N .

When the data were converted into difference scores, a glaring discrepancy appeared between the two experiments: the first one yielded no significant F -ratios in the analysis of variance; while the second, with the single exception of the perception analysis, all of the F 's were significant. This, too, could be the result of a larger N in the second experiment, though another factor that undoubtedly influenced the size of the F -ratios was the presence of a control group in the second experiment. In every analysis, the largest differences were between this one group and all the others. With no control group in the analysis, the F -tests in the first experiment were handicapped not only by a smaller N , but by loss of the greatest contributor to 'between-groups' variance.

CONVERGENT RESULTS FROM THE TWO EXPERIMENTS

It might be fruitful to accept tentatively the above explanation of the lack of significance of 'between-groups' differences in gain

scores obtained in the first experiment, and to compare the direction of differences obtained in the two experiments wherever these are shown to be significant in the second experiment.

The second experiment produced a significant difference in favor of Group I over Group II and of Group I over Group III in every analysis but the one on perception; differences in the first experiment were also in this same direction. No significant differences appeared between Group II and Group III in either experiment.

In the second experiment Groups II and III were significantly superior to Group IV in every analysis but the one on perception. There was no control group in the first experiment, but there is every reason to believe that such a group would have gained even less during the training period than did the control group actually used in the second experiment. The former would have had no experience whatever with the forty experimental combinations¹¹; while the latter did encounter some of them in a haphazard manner during the time that the experimental groups were being trained.

Thus, in every case where the second experiment produced a significant difference in a given direction, a difference in the same direction was found to have emerged also from the first experiment.

DISCUSSION

The statement was made as an introduction to this paper that its major objectives would be the testing of two hypotheses: (1) that tachistoscopic training methods are effective in the teaching of number skills, and (2) that they are more effective than a conventional workbook method.

Actually, there was implicit in the design of the experiments a more precise statement of the second hypothesis; the development of an original tachistoscopic method for comparison with the workbook method, but also with the Willey-Billington method, implied the further hypothesis that this new tachistoscopic technique would prove superior to the established one. This hypothesis was based on a conviction that a period of striving for recall—the 'challenge period', the writer calls it—was important in learning and was missing from the procedure described by Willey and Billington. Their method, on the other hand, was expected to prove superior to the workbook approach because it was designed to rid

¹¹ The course of study did not call for these particular items until later in the year.

subjects of their "uneconomical and disorganized habits of calculation in favor of more efficient direct responses in the form of a continuous emergence of precise, economical, mature patterns of learning." (14) Hypothesis 2, then, really involves a ranking of the three experimental methods of training with reference to their effectiveness in producing improvement in number skills.

Also implicit in the design of the experiment is a set of hypotheses about the relationship of computational skills to perception: (1) subjects who are superior to their fellows in perceptual ability are superior also in computational skill, because the latter is to some extent dependent on the former; (2) improvement in perceptual ability is accompanied by an increased effectiveness in computation; and (3) the tachistoscopic training used in these experiments is effective in developing perceptual ability of the type tested.

The entire hypothetical structure, together with pertinent results from both experiments, is summarized in the following outline:

PERCEPTION

1) "Subjects who are superior in perceptual ability are also superior in number skills." The validity of this hypothesis can be neither confirmed nor rejected on the basis of correlation coefficients obtained before training between perception test scores and arithmetic test scores: in the first experiment, the correlation was insignificant by inspection of the scatterplot, while the second produced a correlation of .30 which is significant beyond the .05 level of confidence. The insignificant r from the first experiment may be due in some measure to the smallness of the N .

2) "Improvement of perceptual ability results in improvement of number skills." This hypothesis could not be tested, primarily because of the results reported under hypothesis 3, below.

3) The tachistoscopic training used in these experiments is effective in developing perceptual ability of the type tested." This hypothesis must be rejected on the basis of the analysis of variance of perception test difference score: F -ratios in both experiments were insignificant, and since the lowest difference score was in each case actually a loss of proficiency, it can be safely concluded that no significant gain was made by any of the seven groups tested.

NUMBER SKILLS

1) "Tachistoscopic training is effective in the teaching of computational skills." This hypothesis can clearly be accepted. In the first experiment, which lacked a control group, all of the experimental groups made significant gains in arithmetic test scores of the four types analyzed; in the second experiment, there was in every analysis a significant difference between the gain made by the control group and the smallest of the gains made by the experimentals.

2) "The three experimental methods can be ranked according to their relative effectiveness in the following order: III, II, I." This hypothesis must be rejected. In the first experiment, the analysis of variance did not disclose significant differences between the groups; but the N was small, and all differences were in the same direction as the significant differences obtained with a larger N in the second experiment. Both experiments ranked the methods as follows: Group I (workbook) first, Groups II and III (tachistoscopic) second; Group IV (control) was last in every analysis.

Caution must attend the extension of these interpretations to situations other than the ones described here. Previous experiments with subjects at the primary-school level, for example, have indicated that the tachistoscope can be of value to teachers in a number of different areas, including arithmetic.

It should be noted also that a good many interaction effects are possible which could not be examined in the present study. Controls were attempted of all the variables that seemed at the outset most likely to influence the results; but on the one hand it is impossible to control every factor that could be included, and on the other hand variance contributed by these factors might prove interesting in its own right. Had conditions permitted, a multivariate design would have been more rewarding.

The results might be challenged from another point of view: it could be objected that the standardization that was necessary for a clear differentiation of methods had the effect of rendering them all stilted and artificial. The criticism is well-taken; it is undoubtedly true that each of the three experimental methods suffered to some extent from the limitations imposed by experimental controls. On the other hand, it is also true that the methods were all carefully planned to get the most out of each within these limitations, and there is no apparent reason why the enrichment that

would be possible in a non-experimental situation should affect one method any more than another.

SUMMARY AND CONCLUSIONS

Previous investigators of tachistoscopic training have not come to any agreement regarding its effectiveness with learners of varying ages and backgrounds working with many different materials. Effects are apparently specific to particular constellations of these factors, and conclusions must be limited to the one constellation which characterizes a particular investigation. The present study is concerned with the effects of fourteen 8-min. training sessions on fourth-grade, public-school children working only with arithmetic combinations (multiplication and division). The writer concludes that in this situation tachistoscopic training is effective in the teaching of number skills but that it is not more effective than an ordinary workbook method of practice—in fact, the two tachistoscopic methods used proved to be actually less effective than the workbook method. Tests of rapid perception showed no gain in any of the groups over the period of training.

REFERENCES

- 1) Catherine Aiken. *Method of Mind-training: Concentrated Attention and Memory*. New York: American Book Company, 1895.
- 2) K. M. Dallenbach. "The effect of practice on visual apprehension in school children." *Journal of Educational Psychology*, 1914, 5, 321-324 and 387-404.
- 3) ——— "The effect of practice upon visual apprehension in the feeble-minded." *Journal of Educational Psychology*, 1919, 10, 61-82.
- 4) Cecil Max Freeburne. "The influence of perceptual span and perceptual speed upon reading ability." *Journal of Educational Psychology*, 1949, 40, 321-352.
- 5) Cyril J. Hoyt, Clayton L. Stunkard, W. Reed Hastie, Clifton A. Gayne, Mildred M. Page, and Paul R. Wendt. "A comparison of two methods of instruction in beginning drawing." *Journal of Experimental Education*, 1952, 20, 265-279.
- 6) Fannie H. Melcer and Brown. "Tachistoscopic training in the first grade." *Optometric Weekly*, 1945, 36, 1217-1219.
- 7) Leslie E. Neville, Editor. "Split-second recognition: A new allied weapon." *Aviation*, 1943, 42, 119, 382, 385.
- 8) Samuel Renshaw. "The influence of tachistoscopic training at far point on the scholastic achievement of first-grade children." *Psychological Optics*, 1945a, 6, 1-21.
- 9) ——— "The visual perception and reproduction of forms by tachistoscopic methods." *Journal of Psychology*, 1945, 20, 217-232.

10) — —. "The Gary tachistoscopic training results five years later." *Psychological Optics*, 1951, 12, 47-49.

11) Hoyt L. Sherman. *Drawing by Seeing*. New York: Hinds, Hayden and Eldredge, 1947.

12) J. Sutherland. "The relationship between perceptual span and rate of reading." *Journal of Educational Psychology*, 1946, 37, 373-380.

13) C. O. Weber. "The acquisition and retention of reading skills by college freshmen." *Journal of Educational Psychology*, 1939, 30, 453-460.

14) R. DeVerl Willey and Lillian Billington. "Arithmetic learning by use of the tachistoscope." *The Keystone Tachistoscope—Advanced Teaching Techniques*. Meadville, Pa.: Keystone View Company, 1951.

15) Fred E. Winger. *Tachistoscopic Training for Beginning Typewriting Instruction*. Unpublished Doctoral dissertation, Oregon State College, 1951.

PERSONAL VALUES AND THE EMPATHIC RESPONSE: THEIR INTER-RELATIONSHIPS

GLENN R. HAWKES

The Iowa State College

and

ROBERT L. EGBERT

Utah State College

Recent researches in the field of behavior have high-lighted the concepts of empathy and personal values. These areas appear to hold promise for students of behavior. The ability to empathize is important in social intercourse. It appears also to be a fruitful source of understanding individual differences in certain areas of personality.

Personal values have shed light upon individual differences in motivation. Knowledge of personal values is making a contribution in such fields as counseling where an understanding of conditions of living desired by the individual is important. In these areas, values form a triad with aptitude and interests.

The purpose of this research has been to explore somewhat the relationship between empathy and personal values. It is felt that through continued research in these areas human behavior may become less and less an enigma. The postulates of George H. Mead (5) have been very fruitful in throwing light on the functioning and structuring of the mind and the self and their relation to society. The 'generalized other' of which Mead speaks has given many hints in the direction of a fuller understanding of the nature of social intercourse. Particularly when Cottrell's rôle theory (1) is added, one gains much insight in taking the rôle of the 'other'. The 'other' appears to be the key to unlock many dilemmas in human relationships. Dymond (2) has made use of the 'empathic response' in her research in the field of "taking the rôle of the other". She has completed the shift which has been developing in the use of the term empathy. In the use to which she has put it, there is no reference to the sensations of inanimate objects being internalized but rather reference is made to the internalizing of

the affect and cognition of the 'other'. Dymond defines empathy as follows: "Empathy denotes the imaginative transposing of oneself into the thinking, feeling and acting of another and so structuring the world as he does."

Woodruff (6), using the value concept as a basis, has developed a theory of motivation which holds much promise. He defines personal values as generalized statements of conditions of living which the individual believes have an effect on his well-being—positive—negative—neutral. The measurement of values under his direction has added to the fields where psychometrics attempts to select and predict. Egbert (3) has refined the measurement of values to a point where objectivity is possible on the part of the psychometrist and his scale has sufficient statistical validity and reliability for survey purposes.

Dymond describes her Rating Test as follows: "The test was made up of four parts, each containing the same six items. In the first part the individual was asked to rate himself, on a five-point scale, on each of six characteristics. In the second part he was asked to rate the other individual as he believed this other would rate himself. In the third part the individual rates the other person as he thinks the other person would rate himself. In the fourth he must rate himself as he thinks the other would rate him. In other words, if two individuals A and B are being tested for their empathy with each other, the procedure would be as follows:

"A. Part 1. A rates himself, (A). Part 2. A rates B as he (A) sees him. Part 3. A rates B as he thinks B would rate himself. Part 4. A rates himself (A) as he thinks B would rate him.

"B. Part 1. B rates himself, (B). Part 2. B rates A as he (B) sees him. Part 3. B rates A as he thinks A would rate himself. Part 4. B rates himself (B) as he thinks A would rate him.

"Therefore a measure of A's empathic ability can be derived by calculating how closely his predictions of B's rating, (A 3 and A 4), correspond with B's actual ratings (B1 and B2). Similarly a measure of B's empathy with A can be obtained by calculating how closely his predictions of A's ratings, (B3 and B4), correspond to A's actual ratings (A1 and A2).

"The six traits which were used as the items in all four parts of the test were: 1. self-confidence, 2. superior-inferior, 3. selfish-unselfish, 4. friendly-unfriendly, 5. leader-follower, 6. sense of humor.

"Although the usual objections to such trait-ratings were recognized, this procedure was followed none-the-less because the ratings were not being used to determine the personality of the subjects nor to determine how accurate the others were to their estimation of this. The test was designed to answer the question how well can the subject transpose himself into the thinking, feeling and acting of the others. If he can do this he should be able to predict how the others will behave in certain defined situations. The situation chosen to test this ability was the subject's ability to predict how others will rate themselves and how they will rate him on these six traits. (A high score indicated low empathic ability and a low score the inverse.)"

For the purpose of this study an adapted form of Dymond's test was devised. This form may be used with an I.B.M. answer sheet.

Egbert's "A Study of Choices Form VII" gives fifteen values for a personal value pattern. He lists and describes the values as follows: (3).

- 1) Esthetic Appreciation—Appreciation of beauty and quality of surroundings.
- 2) Comfort and Relaxation—Primarily physical comfort with some reference to 'mental comfort'.
- 3) Excitement—Experiences which are unusual and which give a thrill.
- 4) Freedom—Lack of physical, mental or moral restraint from any external source.
- 5) Friendship—Personal attachments to people outside the home.
- 6) Family life—The husband, wife and children with the inter-relationships that exist among them.
- 7) Intellectual Activity—Mental activity as an end in itself.
- 8) Personal Improvement—Improvement of appearance and facility with the social graces.
- 9) Power and Control—Control over associates and followers.
- 10) Recognition—Prestige as such, apart, in-so-far as possible.
- 11) Religion—Formalized religion with particular emphasis on belief in a Supreme Being and a life after death.
- 12) Privacy—Being alone and having personal belongings which can be kept from other people.
- 13) Security—Financial and physical security.

14) Society Life—Mingling with the social upper class and being active in clubs and night life.

15) Social Service—Doing things for other people: The Golden Rule.

The device is based on a ranking method. Egbert states there are thirty groups of five statements of values. The problem is to rank the statements in each group according to their importance for the happiness of the respondent. Each value appears ten times in the test and each value is in the same group with each other value at least twice and not more than four times.

The test correlates well with Woodruff's Study of Choices. Informal interview has also indicated its validity. Reliability coefficients based on the Froelich index of test reliability have a median coefficient of 0.690.

Eighty college sophomores and juniors, students of elementary educational psychology, were selected for subjects. Egbert's value test was administered to derive a value pattern for each subject. A questionnaire was filled out by each subject in which he ranked the other subjects from 'know casually in class' to 'close friends'. Four subjects were chosen for a grouping for the administration of the empathy test. The subjects were selected to be in the same group only if their degree of friendship was more than 'casually in class' and less than 'close friends'.

Dymond's empathy test was administered to the population and scored according to the deviation score. This gives each subject an 'empathic ability' score and an 'ability to be empathized with' score.

Table 1 is concerned with the twenty subjects from the population who were lowest and the twenty subjects who were highest for each value pattern concerned. Their empathic scores are also shown in this table.

A difference significant only at the .10 level was found between the low social service and the high social service groups. This difference favors the subjects having the greatest empathic ability. The group nature and social intercourse supposition of social service would be in favor of those having the greatest empathic ability.

The value tests were tabulated for deviation; that is, a score was derived for each individual depending upon his value deviations from the theoretical mean of 30 for each value. Either a positive or

TABLE 1.—EMPATHIC ABILITY FOR GROUPS HIGH AND LOW FOR EACH VALUE

Values	Total	High	Low	Diff between High and Low	t	Level of Signif.
1) Esthetic apprec.	23.72	25.60	23.40	2.20		
2) Comfort	23.72	25.30	22.95	2.35		
3) Excitement	23.72	23.35	22.55	0.80		
4) Freedom	23.72	24.35	23.80	0.55		
5) Friendship	23.72	24.10	24.65	0.55		
6) Family life	23.72	21.10	24.50	2.90		
7) Intellectual act.	23.72	22.90	23.95	1.05		
8) Personal improv.	23.72	23.30	23.30	0.00		
9) Power and control	23.72	24.75	23.90	0.85		
10) Recognition	23.72	23.10	24.50	1.40		
11) Religion	23.72	23.75	23.75	0.00		
12) Privacy	23.72	23.70	24.20	0.50		
13) Security	23.72	23.55	25.20	1.65		
14) Society life	23.72	24.00	25.20	1.20		
15) Social service	23.72	21.45	24.55	3.10	1.89	.10
16) Deviation score	23.72	23.70	25.85	2.15		
	N-80	N-20	N-20			

negative deviation was added as it was felt the direction of deviation was not the significant factor. The means for the high twenty and the low twenty in empathic ability were tabulated and then tested for statistical significance. A 't' score significant at the .05 level was yielded. This indicates a statistically significant greater amount of deviation was found to exist for those subjects with greater empathic ability.

It was felt that the deviation score may be a measure of maturity (4)—the more mature individual having a more defined pattern. He has differentiated his values, precluding many contradictions that might appear in the value pattern of a less mature person. He therefore may need less effort to determine his own status and more effort can be expended to clarify relationships.

Table 2 is the companion of the preceding table. It indicates the mean ability to be empathized with for the high twenty and low twenty for each value named. Significant differences were found to favor family life at the .02 level for the higher group. The value of security was found significant at the .05 level. Again, the high groups had the higher mean. Social service value yielded a 't' of 2.86 which was significant at the .01 level. The high empathic

TABLE 2.—“ABILITY TO BE EMPATHIZED WITH” FOR GROUPS HIGH AND LOW FOR EACH VALUE

	Total	High	Low	Diff between High and Low	t	level
1) Esthetic apprec.	23.83	25.70	22.55	3.15		
2) Comfort	23.83	25.95	23.60	2.35		
3) Excitement	23.83	25.20	22.70	2.50		
4) Freedom	23.83	25.60	24.05	1.55		
5) Friendship	23.83	24.30	25.05	0.75		
6) Family life	23.83	22.20	27.40	5.20	2.77	.02
7) Intellectual act.	23.83	23.70	25.90	2.20		
8) Personal improv.	23.83	24.05	23.35	0.70		
9) Power and control	23.83	23.35	23.95	0.60		
10) Recognition	23.83	24.80	23.35	1.45		
11) Religion	23.83	22.40	26.30	3.90	1.99	.10
12) Privacy	23.83	25.40	23.10	2.30		
13) Security	23.83	21.35	25.60	4.25	2.51	.05
14) Society life	23.83	26.50	23.30	3.20		
15) Social service	23.83	21.75	26.05	4.30	2.54	.02
16) Deviation score	23.83	23.85	28.45	4.60	2.86	.01
	N-80	N-20	N-20			

ability group had the highest mean. Religion was found to be significant at only the .10 level for the high group. These values are all important in group living. Those individuals who function well in social relations appear to also lead in values of a social nature.

The table also indicates the significant differences in the deviation scores. The high group had a greater deviation score, significant at the .01 level, reflecting what may be thought of as greater maturity.

TABLE 3.—VALUE PATTERNS FOR TEN SUBJECTS WITH WHOM OTHERS EMPATHIZE BEST

Values		Values	
1) Family Life	41.40	9) Power and Control	28.40
2) Friendship	38.30	10) Recognition	27.60
3) Security	37.50	11) Intellectual Activity	26.80
4) Personal Improvement	34.60	12) Society Life	23.40
5) Social Service	34.40	13) Excitement	21.70
6) Freedom	33.00	14) Privacy	21.70
7) Religion	33.00	15) Esthetic Appreciation	19.30
8) Comfort	28.90	16) Deviation Score	100.70

The composite value pattern for the ten subjects with whom others empathize best is shown in Table 3. It is noted that the values that rank highest are, in general, those which are heavily weighted by social intercourse.

The correlation between empathic ability and the ability to be empathized with was found to be $+0.335$. Research by Dymond¹ has indicated the finding of a correlation of $+0.60$ between empathic ability and ability to be empathized with.

SUMMARY AND CONCLUSIONS

Dymond's Rating Test for measuring empathic ability and Egbert's Study of Choices, Form VII were administered to eighty college students of educational psychology. They were grouped in diverse ways to reflect certain value pattern combinations.

Results were tabulated statistically to determine degree and type of relationship existing between personal values and empathy.

Results indicate that the subjects with high empathic ability—particularly the 'ability to be empathized with'—tend to have the highest values in the areas where group interaction and social intercourse are major factors. Where the high values are less dependent upon group life and can be satisfied by individual action, empathy scores appear to be lower.

Conclusions that certain other factors are operating in the relationship appear to be warranted. They are:

1) Empathy appears to be related to the amount of dispersion within a given pattern; i.e., if the value pattern is flat, or has few peaks, empathy scores are lower.

2) Empathy seems to be positively related to such value factors as family life and social service. For the values comfort, power and control, and society life an inverse relationship is indicated.

3) A correlation coefficient of 0.335 was found between empathic ability and the 'ability to be empathized with'.

REFERENCES

- 1) L. S. Cottrell. "The analysis of situational fields in social psychology". *American Sociology Review*, 1942, VII, 370-382.
- 2) Rosalind Dymond. "A scale for the measurement of empathic ability". *Journal of Consulting Psychology*, 1949, XIII, 127-133.
- 3) Robert L. Egbert. *The Effect of Some Childhood and Adolescent Ex-*

¹Reported in a Seminar in Social Psychology, Cornell University, Ithaca, New York, 1949.

periences on the Emergence of Values. Unpublished manuscript, Ithaca, New York. Graduate School, Cornell University, 1949.

4) Glenn R. Hawkes. "Use of the MMPI in screening college students for counseling purposes". *Journal of Educational Psychology*, 1950, 41, 116-121.

5) George H. Mead. *Mind, Self and Society*. Chicago: University of Chicago Press, 1934, p. 399.

6) A. D. Woodruff. *Psychology of Teaching*. New York: Longmans, Green and Co., Inc., 1951.

COLLEGE GRADES AND SELF-ESTIMATES OF INTELLIGENCE

ORVILLE G. BRIM JR.

University of Wisconsin

This paper reports a study of the relationship between individual self-estimates of intellectual rank and college academic performance as measured by individual grade-point average. The study continues some of the exploratory work on the influence of non-intellective factors upon high levels of academic and occupational achievement which began in another research program.¹

The major proposition of the study is that the grade-level attained by an individual varies directly with his own estimate of his intelligence. The relevant data are those showing the relationship between the subject's actual grade-point average and self-estimate of intelligence.

It is further proposed that this relationship results in part from a tendency of the grade-point average sought by an individual, as well as the strength of his motivation to achieve it, to vary directly with his own estimate of his intellectual rank. Thus two types of relevant data here are the relationship of responses to questions about grade aspiration and about motivation to self-estimates of intelligence.

METHOD

Questionnaires were administered to one hundred and three students in an advanced social psychology class (no freshmen, eleven sophomores) at the University of Wisconsin.

Data on self-estimated intelligence were collected in the following way. Each respondent was asked to assume that he was among one hundred students taken at random from his complete class at the University, and that this group of one hundred was then given a series of intelligence tests and ranked from 1 (high) to 100 (low) on general intelligence on the basis of the test results. The respond-

¹ The writer was formerly a staff member of the project, "Cultural Factors in Talent Development," under the general direction of Fred L. Strodtbeck of Yale University.

ent was then asked to estimate what his position or rank would be among these one hundred students.

Respondents were given the choice of signing their names or of leaving the questionnaire anonymous. Of the one hundred and three respondents, seventeen did not identify themselves. There were no differences between signers and non-signers on the characteristics of sex and of class in the University. The non-signers' self-estimates of intelligence were higher than those of the signers, the mean of the estimates being the 31st rank in 100 for the former and the 39th rank for the latter. A *t* test, however, (2, Ch. 12), shows the probability of this difference to exceed the .05 level, and therefore in subsequent analysis the two groups are considered together where appropriate.

For the respondents who identified themselves it was possible to get both the cumulative grade-point average and a measure of their intelligence from University records.² The intelligence measure was the American Council Psychological Examination (1948 edition). The A.C.E. test norms used in computing the percentiles were those for the entering class of the individual in question.

RESULTS

1) *Relationship of Actual Grade-point Average to Estimated Intellectual Rank.*—In our analysis it is necessary to separate the influence upon grade-point average of both actual and estimated intelligence. The correlation³ between these two intelligence measures was .43. The subjects in general overestimated their intellectual rank. The mean estimated rank for identified subjects, as noted above, was the 39th. Their mean rank on the A.C.E. was the 52nd percentile, which when reversed in direction to be comparable to our data becomes the 48th. Hence the subjects on the average estimated their rank to be nine percentiles higher than the test results.

The relationship of both actual and estimated intellectual rank to grade-point average is shown in simple form in Table 1. The cell values were generated by cutting the rank order of both actual intelligence and estimates at the median, so as to most equalize the

² The writer is indebted to Dr. Joseph L. Lins, Director of Student Personnel and Studies at Wisconsin, for his assistance in this task.

³ Pearsonian correlation coefficients are used throughout. See McNemar (2, Ch. 6.)

TABLE 1.—MEAN GRADE-POINT AVERAGE BY A.C.E. AND ESTIMATED INTELLECTUAL RANK*

	Estimated Rank			
	Low		High	
	A.C.E. Rank		A.C.E. Rank	
	Low	High	Low	High
Mean Grade-Point Average	1.54	1.67	1.76	1.97

* In terms of letter grades, 1 equals C, 2 equals B, 3 equals A.

distribution, and then computing the mean of the grade-point average of the individuals in the cells.

The differences between means show the influence of actual intelligence, which is to be expected. But the differences also show that with actual intelligence roughly constant, those persons who believe they have higher intelligence get the higher grades.

The correlations between the subjects' grade-point average and both their actual and estimated intelligence are .31 and .32, respectively, which are significant at the .01 level.

With the linear effects of actual intelligence held constant, the partial correlation (3, Ch. 9) between grade-point average and estimated intelligence is .20. The partial correlation between actual intelligence and grades with estimates held constant is also .20. A value of .22 would be necessary to reach the .05 level of significance.

Our interpretation is that both variables, actual and estimated intellectual rank, are making a small but independent contribution to differences in grade-point average, with individuals presumably being influenced predominantly by one or the other.

The multiple correlation of both measures of intelligence with grade-point average was .38, indicating that either of the simple correlations is almost as good a predictive measure as are both variables taken together. None of the correlations, however, is high enough to be of much value in prediction.⁴

2) *Relationship of Desired Grade-point Average to Estimated Intellectual Rank.*—Our predictions were that a high estimate of intellectual rank would be associated with a high grade level sought by

⁴ In addition to the correlation data, a simple Guttman coefficient of predictability was computed for actual and estimated rank combined. In this case C equalled .39.

the student, and a strong motivation to achieve this level. Such predictions would follow from the theoretical work of Horney (1), Murphy (4), Sherif and Cantril (5) and others.

An individual's self-image is composed of segments, e.g., one thinks of himself as big, strong, and handsome. Where segments of the self-image are valued by the individual, ego-motives arise. These motives are directed to goals which, if achieved, the individual can interpret as evidence that he is in fact what he considers himself to be.

The segment of self-image involved in this study is one's conception of himself as 'bright', 'average', 'dumb', and the like. Most normal individuals in our culture positively value such conceptions as brightness; i.e. conceptions of being highly intelligent. Individuals who have such self-conceptions of high intelligence are motivated to protect this conception, to act in such a way that corroborating evidence results.

Among American college students, the achievement of high grades is customarily held to be evidence of high intelligence. Individuals among this group who conceive themselves as ranking high in intelligence would therefore normally seek high grades as evidence for this conception.

However, if we hold that grades are taken as evidence of one's intellectual abilities, then we must recognize at this point the criticism that the correlations reported above may simply reflect the fact that the self-estimates are based on grades, and do not stand in any causal relationship to them at all. Our answer is that the relationship between college grades and self-estimates of intelligence is probably a fluid one in which a gradual accommodation is reached between the self-conception and the realities represented by grades. The remainder of the paper presents data which indicate that self-estimates do play a causal rôle in determining grades, and that therefore the correlations reported above represent in part the adjustment of grades to self-conceptions, and not solely the reverse.⁵

⁵ An experimental design giving an adequate answer to this question would be one in which randomly selected groups of entering college students were given information on their intellectual rank, with one group being placed above their actual test scores, another below, and so on. The groups could then be followed through college, and their achieved grades compared with the grades predicted on the basis of customary psychometric techniques.

TABLE 2.—CHANGE IN GRADE ASPIRATION WITH ASSUMED RISE IN INTELLECTUAL RANK

Response	Number of Subjects
Try for better grades	70
Try for less	0
Not change at all	0
No answer	3

The tendency of aspiration level concerning grades to vary with one's belief about his intellectual rank was explored by means of hypothetical situations. Each subject was asked to assume that he had found out that he was much more intelligent than he thought, that his true rank was much higher, and then was asked to indicate whether he would 'try for better grades' than he was trying for now, try for less than now, or 'not change feelings at all about trying for grades'. The distribution of responses support the prediction. These are shown in Table 2.

The same information was collected again, but with each subject asked to assume that he had discovered that actually his intellectual rank was much lower than he thought. The distribution of responses is given in Table 3.

The writer can offer no systematic explanation for the responses of the eighteen subjects who would try for better grades as a result of finding they were much less intelligent than they thought.

In regard to the 'non-changers' in both situations, it seemed probable that these were subjects who fall toward extremes in self-estimates. Those who already have high self-estimates would be unlikely to raise their aspirations on finding they are even more intelligent. Conversely, those subjects with low self-estimates would be unlikely to lower their desires after finding out they are even lower than they thought.

TABLE 3.—CHANGE IN GRADE ASPIRATION WITH ASSUMED FALL IN INTELLECTUAL RANK

Response	Number of Subjects
Try for better grades	18
Try for less	50
Not change at all	27
No answer	8

This possible explanation of lack of change in aspiration was tested by splitting the subjects into high, middle, and low groups in regard to their own estimates of intelligence, cutting at points which most equalized the numbers in the three groups. There was, however, no significant difference between the three groups in the number of non-changers, thus indicating that the explanation lies elsewhere.

3) *Relationship of Strength of Motivation to Estimated Intellectual Rank.*—As a measure of strength of motivation each subject was asked to indicate what his feelings were when he failed to achieve the grade he had set as a goal in any particular instance; i.e., if he failed to do as well as he was trying to do in regard to grades. The indication of concern was made along a nine-point scale ranging from 'indifferent' (0), to 'extremely upset' (9).

Responses to this item were placed into high and low categories of concern by cutting at the point (between 6 and 7) which made most equal the numbers in each group. The number of responses in each category by subjects in the high, middle, and low groups in regard to self-estimated intellectual rank is given in Table 4. A chi square test indicates the difference in distribution of responses to be significant, with $P < .05$.

The results support our contention that subjects with high self-estimates would show greater concern. However, they show in addition that many of the subjects with lowest estimates are also highly concerned. One explanation of this may be that subjects in this group are concerned for quite practical reasons. The scholastic average of this group is low. They are concerned not over threats to the ego-picture, as are the high subjects, but rather over the

TABLE 4.—DEGREE OF CONCERN OVER FAILURE BY ESTIMATED INTELLECTUAL RANK*

Rank	Concern	
	Low	High
High	16	15
Middle	27	7
Low	17	14

* There were seven 'no answers' on the question about amount of concern.

possibility of being put on probation or dropped from school for failure to keep up their averages.

The relationship between motivation and self-estimates was explored in an additional way, in a hypothetical situation. Each subject was asked to indicate how he would feel if he failed to get the grades he would try for after finding out that he was much more intelligent than he thought. The same nine-point scale for indicating responses was used.

The mean concern indicated in response to this hypothetical situation was six on the scale. This was identical with the mean concern expressed by the subjects over their actual failures. This finding was contrary to the prediction that the mean concern would be higher in the hypothetical case.

Identical information was collected in regard to a hypothetical discovery of much lower intelligence. The mean concern indicated was five on the scale. A *t* test indicates this to be significantly lower than the concern expressed over actual failures, with $P < .001$. This difference is in accord with our prediction.

There was no relationship between original self-estimates of rank, and changes in degree of concern following the hypothetical changes in rank. This is in accord with the finding on lack of changes in grade aspiration under the same conditions.

SUMMARY

A questionnaire study was carried out to determine the relationship between self-estimates of intelligence and college grades. A small but significant correlation of .32 was found. When actual intelligence (A.C.E. percentile) is controlled, the correlation drops to .20, which is slightly below the .05 level of significance.

When the subjects were asked to assume that they discovered their intellectual rank was much higher and much lower than they had estimated, their reported aspiration level for grades showed a corresponding increase and decrease.

Significantly more of the subjects in both high and low thirds of the self-estimated intellectual ranking reported a high degree of concern over failure to achieve the grades they desired.

When subjects were asked to estimate their concern over failure after assuming their intellectual rank to be much higher and much lower than they had estimated, there was no difference from exist-

ing concern over failures in the first case, but in the second it dropped significantly lower.

REFERENCES

- 1) K. Horney: *Our Inner Conflicts*. New York: W. W. Norton, 1945.
- 2) Q. McNemar. *Psychological Statistics*. New York: Wiley, 1949.
- 3) A. M. Mood. *Introduction to the Theory of Statistics*. New York: McGraw-Hill, 1950.
- 4) G. Murphy. *Personality*. New York: Harpers, 1947.
- 5) M. Sherif and H. Cantril. *The Psychology of Ego-involvement*. New York: Wiley, 1947.

SPELLING ABILITY AND VOCABULARY LEVEL OF ONE HUNDRED COLLEGE FRESHMEN

FRANK LAYCOCK

Assistant Professor of Education, University of California, Riverside, Calif.

High-school and college English teachers chronically deplore the poor spelling they find in student essays and occasionally suggest that it hampers effective expression. If an idea seems to demand a certain word, the argument goes, but the word is hard to spell, a prudent writer will search for a substitute he knows he can spell. If he chooses a less suitable substitute, clarity loses out to prudence, and the solution must be to improve spelling.

This reasoning permits three hypotheses: (1) the complexity of prose vocabulary correlates positively with spelling ability (so that poor spellers restrict their written vocabulary to words they can spell); (2) vocabulary complexity shows no correlation with spelling ability; and (3) vocabulary complexity correlates negatively with spelling ability. This investigation tested these hypotheses on a group of college students.

PROCEDURE

Spelling.—In a preliminary trial Hudelson's (3) 'Seven S' list yielded thirty-five words suitable for a college-level spelling test. Fifteen additional, harder words came from an unpublished scale by Luther C. Gilbert, making a total of fifty words. A large class containing nearly one hundred and fifty freshmen took the spelling test. Introductory directions explained the situation and invited coöperation; mimeographed sheets provided space for writing personal data and the fifty words; an unhurried, friendly atmosphere dissipated tension. Each of the words in the test was pronounced clearly, read in a sentence, and pronounced again. The whole process for a word was repeated on request; to finish, the words alone were repeated in order.

Students.—In the class tested, one hundred and forty freshmen turned in usable tests. Further selection eliminated: immigrants handicapped in English; students claiming freshman status whose records showed a year or more of college work; students whose

English placement tests (containing the essays to be used for measuring vocabulary) were not available; and students who had already studied English composition at college. There were one hundred and two students left; two more were arbitrarily excluded to provide an even one hundred. These one hundred, from a course which annually attracted a large student group and had no restrictions of admission, were representative of freshmen beginning their second semester of college.

Vocabulary.—To measure written vocabulary, the essays were used which are part of the matriculation examination in English. Every entrant must take the test and is likely to try his best, because it is widely known that failure means separate non-credit instruction and a certain popular stigma. Instructions clearly state that mechanical errors like spelling are penalized and that wise choice of words determines much of the total score. The suggested length of five hundred words is a generous sample on which to base an average.

A separate tally sheet was prepared for each student, with space for pertinent data taken from the essay folder. Horn's (2) 10,000-word list was used to measure complexity in preference to that of Thorndike (4) or Ayres (1) because it was based on school work and letters as well as published reading material. Horn considered derivatives separately, and had a system of weighting and credits which make the final tabulation as accurate as possible. For each word in a student's essay which was among Horn's 10,000 most common words, a score was tallied to indicate the weighted frequency assigned to it by Horn. This weighted score represented the frequency with which the word was met in tallying over 5,000,000 running words. A separate space was reserved for words not in Horn's list and for proper names (arbitrarily ignored by Horn). Words not among Horn's 10,000 are presumably either more difficult or more specialized—and in either case would therefore not be words most commonly used in discussion. However, the topics listed as acceptable for essays (for instance, "Juvenile delinquency as a post-war problem") called for special and perhaps difficult terminology. These uncommon words were tallied separately even though no weighted score was available, because leaving them out would eliminate any indication of either a rich vocabulary or the restrictions spelling might impose.

RESULTS

Spelling.—For each test, the number of words correctly spelled was converted in the usual way to standard scores having a mean of 50 and a σ of 10. The raw scores ranged from 12 to 48 around a mean of 30.7, with $\sigma = 6.8$. Thus the range of standard scores was from 21 to 75. The raw scores were tested for skewness, and found to be sufficiently normal in distribution to warrant conversion to standard scores.

Vocabulary.—Analyzing vocabulary was complicated by the very wide range of credit numbers (from 11 to over 700,000). Walker's (5) suggestion helped: "When the frequency piles up at one end of a distribution, use of a unit which is fine enough to bring out the form of the distribution in that part of the range where the number of cases is large may produce a very long table if the same interval is used throughout. In such cases, the width of interval is sometimes changed from one part of the range to another. Computations may be carried out with the gross scores which correspond to the midpoints of the various intervals."

Accordingly, intervals were arranged, and for each word the midpoint of the interval within which its credit number fell was tallied. Each essay thus reduced to a tally of interval midpoints, and the mean of these midpoints represented the vocabulary level of the essay. For the one hundred essays, scores ranged from 208,706 to 303,891 around a mean of 254,216, with $\sigma = 9,895$. Thus, in an essay given a score of 208,706, the average word was equivalent to one which Horn tallied 208,706 times in examining over 5,000,000 running words. These raw vocabulary scores were converted to standard scores for easy comparison with spelling standard scores. However, a reversal was necessary. Large spelling scores represent superior ability; large vocabulary scores indicate use of common words—that is, in this context, a poorer vocabulary. Vocabulary standard scores were therefore reversed by subtracting them from 100 (theoretical maximum) before being entered, so that a large standard score would represent superior vocabulary.

Correlation of spelling and vocabulary scores.—The aim of the investigation was to see how spelling and vocabulary relate to each other. Correlating the two sets of scores gave a Pearson $r = .075 \pm .067$, indicating a very nearly chance relationship. This absence of correlation supports the second hypothesis, and it therefore

TABLE 1.—SPELLING TEST: BEST, MIDDLE, AND WORST STANDARD SCORES WITH CORRESPONDING VOCABULARY STANDARD SCORES

Worst Scores			Middle Scores			Best Scores		
Stud.	Splng.	Vocab.	Stud.	Splng.	Vocab.	Stud.	Splng.	Vocab.
1	21	43	46	50	54*	91	63	60*
2	25	36	47	51	49*	92	63	51
3	28	65	48	51	52*	93	65	62*
4	28	55	49	53	42	94	66	38
5	28	42	50	53	62	95	66	35
6	28	53	51	53	59	96	66	45
7	30	57	52	54	47	97	66	43
8	31	50	53	54	44	98	68	64*
9	31	37	54	54	58*	99	68	66*
10	31	34*	55	54	49*	100	75	58

* Vocabulary scores within five standard score points ($= 0.5 \sigma$) of corresponding spelling scores.

cannot be concluded that poor spelling ability noticeably constricts written vocabulary.

Further analyses.—In case the over-all correlation concealed relationships that might be revealed by extreme scores, two comparisons were made: (1) the best ten, worst ten, and middle ten spelling scores were paired with their corresponding vocabulary scores; and (2) the reverse comparison placed the best, worst, and middle vocabulary scores alongside corresponding spelling scores. Tables 1 and 2 show this comparison. The small number of entries in each group did not warrant elaborate statistical procedures. 'Close relationship' was arbitrarily assigned to scores within five standard score points of each other, i.e., within 0.5σ . Using this criterion, there was little discernible relationship: isolating the best, worst, and middle scores did not reveal many closely related pairs.

Another possibility suggested itself. Since the directions offered many topics on which to write essays, perhaps a comparison among students discussing the same topic would show a pattern. Students choosing a common theme may discuss it at different vocabulary levels. Vocabulary level here could be measured by considering the number of words not in Horn's list, and the quality of vocabulary thus indicated compared with spelling ability. Table 3 groups four

TABLE 2.—VOCABULARY TEST: BEST, MIDDLE, AND WORST STANDARD SCORES WITH CORRESPONDING SPELLING STANDARD SCORES

Worst Scores			Middle Scores			Best Scores		
Stud.	Vocab.	Splng.	Stud.	Vocab.	Splng.	Stud.	Vocab.	Splng.
40	23	48	23	49	42	93	82	65*
90	28	63	66	49	57	76	63	58*
22	30	42	34	49	46*	98	64	68*
67	30	57	55	49	54*	33	65	45
84	33	60	47	49	51*	3	65	28
10	34	31*	82	49	60	59	65	55
95	35	66	8	51	31	99	66	68*
43	35	50	20	51	42	16	69	40
2	35	25	61	51	55	75	72	58
58	36	55	69	51	58	73	74	58

* Spelling scores within five standard score points ($= 0.5 \sigma$) of corresponding vocabulary scores.

TABLE 3.—SPELLING STANDARD SCORES AND NUMBER OF WORDS IN ESSAYS
NOT IN HORN'S LIST, OF SELECTED STUDENTS GROUPED ACCORDING
TO FOUR ESSAY TOPICS

[illegible]

topics which together accounted for half the essays. Within each group, students are arranged according to spelling score, with the corresponding vocabulary score entered to show the number of words which are not in Horn's list. Here again no close relationship appears.

These results apply to the restricted situation investigated. Other populations, however, and other ways of measuring spelling and vocabulary may show different patterns. Intelligence or reading or motivation may influence spelling and vocabulary in ways separate from any mutual relationship between these two factors. Special fields at the college level, like mathematics, English, or foreign languages, may have particular characteristics.

CONCLUSIONS

1) College freshmen vary in spelling ability and written vocabulary. Spelling variation in the group tested was similar to that in the general population, although about a higher mean. Vocabulary varied, too, but there are no general population data for comparison.

2) Of the three hypotheses set up for testing, only the second is justified: vocabulary complexity shows no correlation with spelling ability. There is no evidence in this study that freshmen inhibit their written vocabulary because of fear of misspelling.

3) Since there is no relationship between spelling and vocabulary, spelling improvement by itself should not be expected to lead to better or freer written vocabulary.

SUMMARY

In this study a spelling test screened out one hundred university freshmen who had not studied English composition or spelling beyond high school, and whose knowledge of English was not affected by foreign extraction or unusual experience. Their spelling scores were distributed normally from high to low. Each of these freshmen had written as part of the entrance examination in English a 500-word essay, and had been warned that choice of words and clarity of expression were important and that mechanical errors like misspellings would be penalized. Scores were assigned to each essay to indicate the average frequency of appearance of the words used, based on Horn's (2) list of 10,000 English words most commonly used in writing. Each student's spelling and vocabulary

scores were compared. Both Pearson r ($= .075$) and inspection of selected cases found only a chance relationship between spelling ability and written vocabulary level. The methods used with this group of one hundred freshmen show that low spelling ability does not noticeably hamper vocabulary choice when there is pressure to write as well as possible. Examining spelling and vocabulary in other settings and with different instruments should help to test this conclusion.

REFERENCES

- 1) L. P. Ayres, *The Spelling Vocabularies of Personal and Business Letters*. New York: Russell Sage Foundation, 1913.
- 2) Ernest Horn, "A basic writing vocabulary: 10,000 words most commonly used in writing." *University of Iowa Monographs in Education*, First Series, No. 4 (April 1, 1926).
- 3) E. Hudelson and Others, "Sixteen spelling scales standardized in sentences for secondary schools" (the *Seven S Study*). *Teachers' College Bulletin* (Columbia University), 1921.
- 4) E. L. Thorndike, *Teacher's Word Book of 10,000 Words*. New York: Bureau of Publications, Teachers' College, Columbia University, 1921.
- 5) Helen M. Walker, *Elementary Statistical Methods*. New York: Holt, 1943. pp. 34; 120-21.

RORSCHACH CONFIGURATIONS ASSOCIATED WITH COLLEGE ACHIEVEMENT¹

CHARLES C. McARTHUR AND STANLEY KING

Department of Hygiene, Harvard University

Vorhaus (5) recently defined four test patterns that she found accounting for nearly three-quarters of the Rorschachs she had collected from children attending a remedial reading center. She was able to explain the children's 'reading problems' as symptoms of long-standing character difficulties revealed by the four test syndromes. The four test patterns that Vorhaus described may be characterized as follows:

Type I is the merely formal record, with no other determinant save the form of the blot being used more than twice.

Type II is the animal-movement dominated record, with nothing else save form being much used.

Type III is the human-movement dominated record, with no accompanying development of other determinants.

Type IV is dominated by inanimate movement and color responses, with no other determinants being much utilized.

We will interpret the psychological meanings of these patterns as they arise in our discussion. Detailed interpretations of them are suggested in Vorhaus' article.

Although in age her subjects ranged from six to twenty-one, and although academic difficulties are not always a result of 'reading problems', we thought it worth while to test her patterns as prognosticators of academic difficulties in college. Past attempts to relate the Rorschach to college adjustment have met more failure than success. In the most hopeful study to date, Munroe (4), using group administration of the test, showed a relation between Rorschach patterns and student success at Sarah Lawrence, but an attempt to apply her method at the University of Chicago (1) brought negative results.

¹ This experiment was supported in part by the Department of Hygiene, Harvard University, the Harvard Medical School Department of Psychiatry and the Department of Social Relations; and in part by Air Force Contract A.F. 33 (038) 201-42, from the School of Aviation Medicine, Randolph Field, Texas.

PROCEDURE

The writers set out to compare the frequency of Vorhaus types among individual Rorschach protocols collected from two groups: (1) one hundred and thirty-seven referrals during the years 1950-1952 to Harvard's Department of Hygiene, all of these students being in some sort of academic or personal difficulty, and all tested by McArthur; and (2) seventy-four controls, who were being tested as part of another experiment at Harvard Medical School. The latter were all Harvard College juniors or seniors in the classes of '52 and '53 and were selected at random from the class lists. They were classmates of, or at least undergraduates contemporaneous with, the Hygiene referrals. They were tested by King.

Our expectations were first that Vorhaus types I and II would be rare in our two samples of college students, inasmuch as they seem to lack the resources to carry them so far in their education. Secondly, we expected that type IV would be common, and would occur more frequently in the Hygiene cases than in the controls. Vorhaus' description of this type sounds like a common variety of college problem-student, known already to the physicians at Harvard. As Vorhaus says:

"The implication . . . is that the subject is responsive to affective stimulation; indeed, he is responsive to a point where moments of strong feeling occur as often, or almost as often, as do those when a more surface pleasantness is all that is evoked The subject may have succeeded in repressing recognition that this is so.

"We again glimpse . . . the 'good' home, the submissive child, and the awareness of pressure Since the resentment (of the pressure) cannot be overcome, the psychological need becomes that of preventing it from being experienced as associated with the environment. This is done by turning the hostility against the self, for the 'guilt' of harboring the resentment With this accomplished, the subject is able to feel that "Mother and Father are entirely just in all their demands and expectations. It is I who am guilty for not coöperating with them. It is because of my inadequacy and inferiority that their 'good' plans have not worked out."

This unhappy state of mind is seen all too commonly in the men who come, or are sent, to the Department of Hygiene for help. They are usually in academic trouble as a matter of administrative fact or else are 'failing' in their own eyes because they have not

met their own and their parents' level of aspiration. In this respect these Hygiene cases are probably quite similar to the reading clinic cases seen by Vorhaus.

Her description of type IV did not seem to fit what was known of the men in the control group. These men had been extensively studied in an experiment on laboratory stress-inducing situations (2) and on the whole had been judged to be stable, well-adjusted students. Therefore, we had grounds for expecting that type IV would be less prevalent among this randomly selected group than among the Hygiene cases.

Before making direct comparisons, it was necessary to be certain that our results were not contaminated by differences in scoring. To test inter-scorer reliability, McArthur rescored twenty of King's protocols, selecting those in which slight changes in scoring would be most likely to result in changes of type. The result was that only one test needed to be reclassified. The question of the effect of the personality of the tester on Rorschach responses could be checked in only one case when both of us tested one subject; in that instance, the configuration was not altered.

RESULTS

When we compared the Hygiene cases with their controls, the results were as predicted. Vorhaus types I, II, and III were comparatively rare, and in no case numerous enough to differentiate significantly between controls and Hygiene cases. Type IV protocols were more frequent, differentiating the Hygiene group from the control group at the .01 confidence level. Furthermore, Table 1 shows that this contrast was reliable over two years of Hygiene experience.

Two circumstances in the data (which are presented in Table 1) raised questions that needed exploration. Only half of the Hygiene cases were included in the four Vorhaus types and only one-quarter of the control cases. While that difference was in the predicted direction, and was itself a satisfactory finding, it leaves fifty to seventy-five per cent of the Rorschachs unexplained. Then, too, the contrast between Hygiene and controls with respect to type III records was, though insignificant, in the wrong direction. What is happening here?

An examination of the Vorhaus check list defining type III shows that its characteristics are: a) M (human movement) has a

TABLE 1.—RELATIVE FREQUENCIES OF RORSCHACH TYPES IN HYGIENE REFERRALS AND CONTROLS

Type	Hygiene '50-'51	Hygiene '51-'52	Controls	Significance
Vorhaus I	7 per cent	2 per cent	7 per cent	none
Vorhaus II	2 per cent	0 per cent	3 per cent	none
Vorhaus III	5 per cent	6 per cent	14 per cent	none
Vorhaus IV	35 per cent	39 per cent	2 per cent	.01
N	= 55	82	74	

numerical value of 3 or more. b) FM (animal movement) cannot exceed one and one-half times M. c) m (inanimate movement) does not exceed one-half times M. d) the color sum (weighted total of responses to color) is less than 3.

On casual inspection, one would not take this as a syndrome that would bode ill for academic performance or for reading skills. Traditionally, the M response has been taken to be the intellectually 'best' kind of response, and type III seems to be an M-dominated record.

The rationale, of course, is that the type III person has great assets, as shown by his movement responses, but that these assets are not put to socialized use. Without adequate development of the color responses (as quantified by condition d), one suspects that the type III man has a rich but too exclusively inner mental life. Vorhaus (5) feels that "the instinctual drives have been integrated into the individual's total functioning, thus becoming the source of creative energy and achievement," but that there is "failure, both in the capacity to establish rapport and in the ability to react to the environment deeply and genuinely." The result, she suggests, is a "determination to preserve his creative drives as his personal (and secret) treasure."

There are those who would say that such a person would be a 'natural' student for Harvard. This is the born resident of the ivory tower, who should fit into a university that has been characterized by F. L. Wells (7) as 'dominantly cerebrotonic'. Yet even 'cerebrotonic' schools demand that students pass tests. Even in such institutions, some 'capacity to establish rapport' is a prerequisite for getting grades. It should therefore be the type III people, with the addition of adequate color response, who make

the best Harvard students. If the color response may be taken as indicating the ability to relate inner resources to outer environment, such modified type III subjects should presumably display resources that are available for socialized use and are not, like the resources of pure type III people, retained as 'private treasures'.

We therefore may define a new type, IIIA, as people who meet the first three criteria of Vorhaus type III but are not bound by type III's fourth limitation. Type IIIA people show: a) M with a value of 3 or more. b) FM less than one and one-half times M. c) m less than one-half times M. d) color sum of 3 or more.

Empirically, they turn out very well. They are found in our groups at the following rates:

	Hygiene '50-'51	Hygiene '51-'52	Controls	Significance
Type IIIA	7 per cent	7 per cent	33 per cent	.01

The difference is in the predicted direction and reaches a good level of confidence. Once again, the Hygiene figures look reliable.

An even stricter definition of the ideal student's Rorschach might demand that his color responses be developed in the direction of form-dominated color responses rather than of the use of color alone or color dominating form as a determinant. That, on orthodox Rorschach theory (3), would indicate not only ability to relate oneself to the environment, but specific skill at tying one's impulses in with the social environment. It is usually FC that is taken as the sign of 'ability to establish rapport'.

So we made a stricter definition of our 'good student' subtype, demanding, beside the criteria for type IIIA, that FC (form-color responses) should numerically exceed CF (color-form) and C (pure color). For this narrower group (which, of course, is selected within the ranks of type IIIA), we find these frequencies:

	Hygiene '50-'51	Hygiene '51-'52	Controls	Significance
Type IIIB	0 per cent	1 per cent	22 per cent	.01

Though we lose cases, the exceptions to our rule are almost eliminated as we sharpen our definition. In fact, the lone Hygiene record showing this type of M and FC dominated scores was made

by a boy who had been referred for purely vocational guidance reasons and not for the usual academic, personal and administrative troubles.

We must hasten to point out that well-developed M and FC responses are usually sufficient but never necessary conditions for academic success. Wells (6) has shown, in a recent study at Harvard, that twenty per cent of the National Scholars here actually had no human movement responses!

From these facts, we may reach certain firm conclusions and suggest other likely hypotheses.

First, we have offered additional evidence for the validity of Vorhaus' typology. Those types for whom she suggests minimal resources are not much seen at Harvard, while her type IV, the only one commonly admitted to this college, is found almost exclusively among people who were in academic difficulty. (We might add that type III, seen at Harvard with some frequency and not characteristic of Hygiene cases, was the type for which Vorhaus foresaw the best prognosis.) The fact that her scheme could be transferred from New York to Cambridge and from pre-college to college students suggests that it possesses more widespread validity than is always found in methods for identifying poor students with the Rorschach.

Second, both Vorhaus' types and our extensions of her typology to include some 'desirable' academic syndromes, are based on usual Rorschach theory. Both typologies have been empirically shown to be valid and reliable. We have therefore some inferential evidence that the Rorschach, interpreted in the customary manner, can lead to correct conclusions about college performance. The limitation seems to be that only some Rorschachs point directly to good or poor academic performance, while other Rorschach patterns occur infrequently or are ambiguous as to their significance.

Third, the positive results shown here point out the need of studies in which the Rorschachs are administered as the student enters college. Ours is a study after the fact. We cannot establish whether the test patterns we observed were long-term characteristics of the men tested or whether the test they gave reflected their recent experience with years or months of college success or failure. The need is to administer the test to a sample of an entering class and immediately make sealed predictions, to be checked after the student has accumulated his good or poor academic record. Our

present findings suggest that such a study would have positive and practically applicable results.

REFERENCES

1) Lee J. Cronbach. "Studies of the group Rorschach in relation to success in the College of the University of Chicago." *J. educ. Psychol.*, 1950, 41, 65-82.

2) S. H. King. *Emotional and Cardiovascular Responses during Stress*. Unpublished doctor's dissertation, Harvard University, 1953.

3) B. Klopfer and D. M. Kelley. *The Rorschach Technique*. New York: World Book Company, 1942

4) Ruth L. Munroe. "Prediction of the adjustment and achievement performance of college students by a modification of the Rorschach method." *Applied Psychol. Monogr.*, No. 7, 1945.

5) Pauline G. Vorhaus. "Rorschach configurations associated with reading disability." *J. Proj. Tech.*, 1952, 16, 3-19.

6) F. L. Wells. "Rorschach and Bernreuter procedures with Harvard National Scholars in the Grant Study." *J. Genet. Psychol.*, 1951, 79, 221-260.

7) — — — "Should This Student Be at Harvard?" in Thurstone, L. L. (Ed.) *Applications of Psychology*. New York: Harper & Brothers, 1952.

A COMPARISON OF THE STANFORD-BINET, REVISED FORM L, AND THE CALIFORNIA TEST OF MENTAL MATURITY (S-FORM)¹

WILLIAM D. SHELDON

Director of the Reading Laboratory, Syracuse University

and

GEORGE MANOLAKES

Ball State Teachers College, Muncie, Ind.

An investigation was made of the characteristics of good and poor readers in grades I-XII in eight school systems in the State of New York. The present article presents the results of the Stanford-Binet Intelligence Scale, Revised Form L, and the California Test of Mental Maturity (S-Form), administered to each child studied in grades I-VI. This study evolved from the need of comparing the two tests in order to assess the intellectual status of the pupils tested in a more accurate manner. The Grace Arthur, Form II, was also administered to each of the pupils in the study and a comparison between this test and the Stanford-Binet has been published.²

SELECTION OF SUBJECTS

The subjects represent ten per cent of all the pupils in grades I-VI in the eight schools participating in the study. The population included 4220 pupils from fourteen elementary and four rural schools. Included in this study were four hundred and twenty-two pupils. They were selected by their teachers in each of the schools. In making their selections, the teachers used three criteria: (1) achievement tests in reading, administered before selection; (2)

¹ Authors wish to acknowledge the coöperation of the Bureau of School Services, School of Education, and the Psychological Services Center of Syracuse University in making this research project possible.

² G. Manolakes and W. D. Sheldon, "A comparison of the Grace Arthur, Revised Form II, and the Stanford-Binet, Revised Form L", *Journal of Educational and Psychological Measurement*, 12, Spring 1952, pp. 105-108.

their own rating of each pupil's status in reading; and (3) test scores derived from intelligence tests administered before the selection. The teachers were asked to choose from their classes, five per cent of their pupils who were good readers and five per cent who were poor readers. The four hundred and twenty-two pupils tested ranged in ages from five years ten months to twelve years eleven months.

PROCEDURE

The Stanford-Binet and the California Test of Mental Maturity were administered to each of the four hundred and twenty-two pupils in the study by members of the research team. The conditions under which the tests were administered were the best that could be obtained in the various schools. The Stanford-Binet test was administered individually by skilled clinicians while the California Test of Mental Maturity was administered to small groups by individuals trained in giving this instrument. The time interval between the two tests ranged from one day to four months, with an average interval of four weeks.

RESULTS

Table I reveals that there is no significant difference between the means of the California Test of Mental Maturity and the Binet. This difference was measured by both the t test and by critical ratio of the difference and its standard error. A coefficient of correlation of .702 was obtained from a comparison of the distributions. The stability of this statistic was determined by Fisher's r to Z transformation; and at the .01 level, this correlation coefficient would range .629 to .757.

Table I also reveals that the Binet distribution is skewed significantly. This might be expected due to the manner in which this group was selected. However, the results of the California Test of Mental Maturity were not significantly skewed. The measure of kurtosis for both distributions cannot be considered significantly different from normal.

According to Table II, 44.6 per cent of the scores obtained from the same individuals differed greater than 10 IQ points. Twenty-four and six-tenths per cent showed a difference greater than 15 IQ points; while 13.7 per cent of the cases showed differences greater than 20 points.

TABLE 1.—FREQUENCY DISTRIBUTION AND CONSTANTS FOR CALIFORNIA
TEST OF MENTAL MATURITY (S-FORM) AND STANFORD-BINET,
REVISED FORM L

Binet	IQ	CTMM
6	160-169	0
11	150-159	4
11	140-149	10
28	130-139	35
61	120-129	60
66	110-119	68
86	100-109	91
76	90-99	95
59	80-89	33
14	70-79	12
4	60-69	7
0	50-59	0
0	40-49	1
422	M	422
108.443	Mean	107.806
.925	s_m	.819
s_{m1-m2}		.680
CR		.937
P		.347
106.244	Median	106.423
18.984	S.D.	16.804
.653	s_s	.578
s_{s1-s2}		.623
CR		3.497
P		.0002
66-169	Range	47-156
28.135	Q_U-Q_L	25.881
Skewness		
.491	g_1	.015
.119	s_{g1}	.119
.00002	P	.4433
Kurtosis		
.0396	g_2	-.0136
.238	s_{g2}	.238
.4325	P	.4761

Correlation

R	.702
s_r	.0247
z	.867
s_z	.049
$z \pm 2.586_s$	.741 to .993
r range	.629 to .757

TABLE II.—DIFFERENCES IN IQ POINTS BETWEEN THE STANFORD-BINET, REVISED FORM L AND THE CALIFORNIA TEST OF MENTAL MATURITY (S-FORM)

Binet Higher		IQ Difference	CTMM Higher	
per cent	N		N	per cent
15.2	64	1-5	65	15.4
11.1	47	6-10	49	11.6
10.0	42	11-15	42	10.0
4.7	20	16-20	26	6.2
3.6	15	21-25	9	2.1
1.7	7	26-30	10	2.4
1.4	6	31-35	4	0.9
0.7	3	36-40	1	0.2
0.5	2	41-45	0	0.0
0.2	1	46-50	0	0.0
49.1	207	No difference 9 cases 2.1 per cent	206	48.8

The average algebraic and arithmetic differences between the test at the chronological age levels are indicated by Table III. The general trend of the average algebraic differences between the scores for individuals indicates that the California Test for Mental Maturity yields scores higher than the Binet at the six- and seven-year level, while the Binet was higher at ages nine, ten, and eleven. The scores were in greatest agreement at the eight-year level. The differences at ages twelve and thirteen are not considered as too indicative or significant because of the small number of students at these age levels.

The average arithmetic differences between the IQ scores indicate an average absolute difference of 10.943 between the two tests. The greatest difference is in the ten-year-old group.

In Table IV a significant trend is revealed by a comparison of the results of the two tests by IQ groups. The average algebraic

TABLE III.—AVERAGE ALGEBRAIC AND ARITHMETIC DIFFERENCES BETWEEN THE STANFORD-BINET, REVISED FORM L, AND THE CALIFORNIA TEST OF MENTAL MATURITY (S-FORM) AT THE CHRONOLOGICAL AGE LEVELS

Age	N	Algebraic Difference*	Arithmetic Difference
6	61	-4.197	11.527
7	82	-1.890	9.415
8	70	0.314	8.788
9	71	3.606	12.535
10	56	3.839	14.446
11	59	3.305	10.763
12	18	-2.778	8.278
13	5	3.000	9.000
Total	422 Average	0.573	10.943

* Negative numbers indicate difference in favor of the California Test of Mental Maturity scores.

differences indicate that at the upper Binet levels, the Binet scores are higher than those of the California Test of Mental Maturity. This difference decreases almost progressively until the average range is reached. At and below the average range, the California Test of Mental Maturity yields a higher IQ.

Regression equations have been computed to demonstrate the disparity between the tests on different IQ levels. The tables further substantiate the differences already indicated. The re-

TABLE IV.—AVERAGE ALGEBRAIC DIFFERENCES BY IQ GROUPS DETERMINED BY BINET IQ

	N	Algebraic Difference*
160-169	6	33.833
150-159	11	18.273
140-149	11	13.909
130-139	28	10.607
120-139	61	2.787
110-119	66	5.712
100-109	86	-2.965
90-99	76	-4.566
80-89	59	-6.475
70-79	14	-9.143
60-69	4	-0.075

* Negative numbers indicate difference in favor of the California Test of Mental Maturity.

gression formula for predicting Binet IQ's from C.T.M.M. IQ's is:

$$X^1 = .793Y + 22.953 \quad s_{x.y} = 13.517$$

The regression formula for predicting C.T.M.M. IQ's from Binet

$$\text{IQ's: } Y^1 = 621X + 40.463 \quad s_{y.x} = 11.964.$$